UNIVERSITY OF LIVERPOOL

BIOPHYSICAL CHARACTERISATION OF SAP97 AND AKAP79

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

MARK DAVID TULLY

JANUARY 2011

Abstract

Molecular scaffolds work to dynamically assemble signal transduction pathways in certain cellular locations at the precise physiological time required. Scaffold proteins often contain multiple binding domains that allow more than one partner at once, helping to assemble a biochemical pathway or protein network. This thesis aims to characterise two such scaffold proteins.

The first is a major scaffold protein, Synapse Associated Protein 97 (SAP97) that facilitates the clustering of potassium and glutamate channels to the post synaptic density. SAP97 contains multiple domains all joined by unique unstructured linker domains that are thought to provide flexibility within the protein and have proved problematic in previous attempts to crystallise the protein.

Three large multiple domain SAP97 constructs, Δ546, Δ461 and Δ70 were expressed and purified. These were then characterised using SAXS, SEC-MALLS, and NMR. SAXS studies demonstrated that SAP97 constructs existed in an ensemble of forms in dynamic equilibrium. Two dominant populations of an extended and compact form were discovered, propagated through the long unstructured linker regions between the PDZ domains. Subsequent NMR confirmed that the PDZ domains were mobile and were not influencing the GK-SH3 domain.

The second scaffold, A-Kinase Anchoring Proteins 79 (AKAP79) that mediates phosphorylation through PKA at the plasma membrane was also characterised using several biophysical techniques. Full length AKAP79 which is notoriously difficult to work with was successfully cloned, expressed and purified and using NMR found to be unstructured. The AKAP79 calmodulin binding domain, a short 20 amino acid peptide with a 1-11 CaM binding motif was also cloned, expressed and purified and was shown to bind with calmodulin in which both their N-and C-domains are involved.

The interactions between SAP97 and AKAP79 were investigated. The central domain of AKAP79 that has previously been shown to interact with SAP97 was cloned, expressed and purified using a novel process of Ligation Independent Cloning. However, central domain was shown not to bind to SAP97 under the conditions used with NMR and ITC.

These new data help to further the understanding of the nature of scaffold proteins SAP97 and AKAP79 and their interactions with other proteins.

Contents

List of figures

List of tables

Acknowledgments

I would like to thank my supervisor Lu-Yun Lian for all the help and support that she has given me throughout my time here. Without her efforts I don't think I would have gotten to the end.

Many thanks must also go to Dr Marie Phelan for all her help with the NMR analysis and using the alien Linux systems; Dr Guenter Grossman for his expertise in SAXS analysis that prevented me from pulling out all my hair and to the rest of Lab C for their support through the long periods were nothing seemed to work.

Finally I would like to thank my family and friends who have pushed me on to meet all the goals I set and have stood by me though all the ups and downs these last years have provided. Special thanks have to go to my wife, Katherine for whom I could not live without. Her love and support have provided the shining beacon that I have followed when things appear at their darkest.

Glossary

AKAP	A Kinase Anchoring Protein
β_1-AR	β-adrenergic receptors
CaM	calmodulin
CaN	calcineurin
CV	column volume
Dlg	discs large tumour suppressor
EDTA.	ethylene-diamine-tetra-acetic acid
FRET	fluorescence resonance energy transfer
Gk or Guk	guanylate kinase
GF	gel filtration
GKAP	guanylate kinase associated protein
GST	glutathione S-transferase
IPTG	isopropyl-beta-D-thiogalactopyranoside
K^+	Potassium ion
Kir	potassium inward rectifier
LB	Luria broth
LIG	Ligation Independent Cloning
MAGUK	membrane associated guanylate kinase containing proteins
NMDA	N-methyl D-aspartate
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline

PDZ	binding domain containing 90 amino acid repeat first found in PSD-95, Drosophila large Disc protein and protein ZO-1
PIP_2	phosphoinositides
PtdIns(4,5)P_2	phosphatidylinositol 4,5-bisphosphate
PKA	type II cAMP-dependant protein kinase
PP	precission protease
PSD95	postsynaptic density 95
R.	PKA regulatory subunits
SAP97.	synapse associated protein 97, a member of the MAGUK Family
SAXS	small angle X-ray scattering
SH2	src homology 2 domain
SH3	src homology 3 domain
SPR	surface plasmon resonance
SUR	sulfonylurea receptor
TEMED.	N,N,N',N'-Tetramethylethylenediamine
TFE	Trifluroethanol
ZO-1	zonal occludens

CHAPTER 1

INTRODUCTION

1.0 Introduction

In the body cells are constantly communicating with each other, passing messages between each other, recognising them, interpreting them and responding to them. These messages from external stimuli are often simple chemicals or more complex hormones and the message carried is relayed from the cell surface to specific intracellular targets via protein – protein interactions that has been likened to a complex biochemical circuit (Harris and Lim, 2001). These signal transduction pathways are still not fully understood but a mechanism to organise and localise all the specific proteins must occur. This is currently thought to be achieved by one of two mechanisms, either bringing together active signalling molecules into a multiple protein network or by activating dormant enzymes (signalling molecules) that are held in position close to their substrates (Pawson and Scott, 1997). The proteins charged with the task of fixing in position and aligning the network are often termed the scaffold.

The work described in this thesis highlights a small part of a protein network that closely follows the second network conditions above. An ion channel, the potassium inward rectifier is held in position and activated by a complex network of protein-protein interactions. The main proteins involved that will be studied in some detail are SAP97, AKAP79, Calmodulin, and KIR2.1. Each of these will be discussed in the following sections along with descriptions of scaffold proteins.

1.0.1 Scaffolding, Adaptor and Anchoring Proteins

When constructing a house a scaffold is required to provide support and to allow the correct tools and materials to be assembled in the right order at the right time. The cellular machinery is no different. Molecular scaffolds work to dynamically assemble signal transduction pathways in certain cellular locations at the precise physiological time required. The scaffold has many functions; it controls the assembly of a multitude of protein and is involved with the trafficking and the subcellular location of protein networks (Welling, 2008).

The post synaptic density (PSD) of a cell is an electron dense area under the plasma membrane that contains receptors, scaffold proteins, signalling proteins such as kinases and phosphatases all linked to the actin cytoskeleton. The distribution, concentration and organisation of specific protein components are regulated by the scaffold proteins (Lisman et al., 2002). This is shown schematically in Figure 1.0 that shows the role PDZ domains play in the pre and post synapse; localising the receptor to the membrane before tethering it and assembling a large molecular complex (Kim and Sheng, 1996; Kim and Sheng, 2004). Figure 1.0 also highlights the complexity of receptor networks that are created and rely upon the scaffold proteins to function.

One example of a scaffold is the protein SAP97 that facilitates the clustering of potassium channels to the post synaptic density (Doyle et al., 1996; Froehner, 1993; Kim and Sheng, 1996; Leonoudakis et al., 2001).

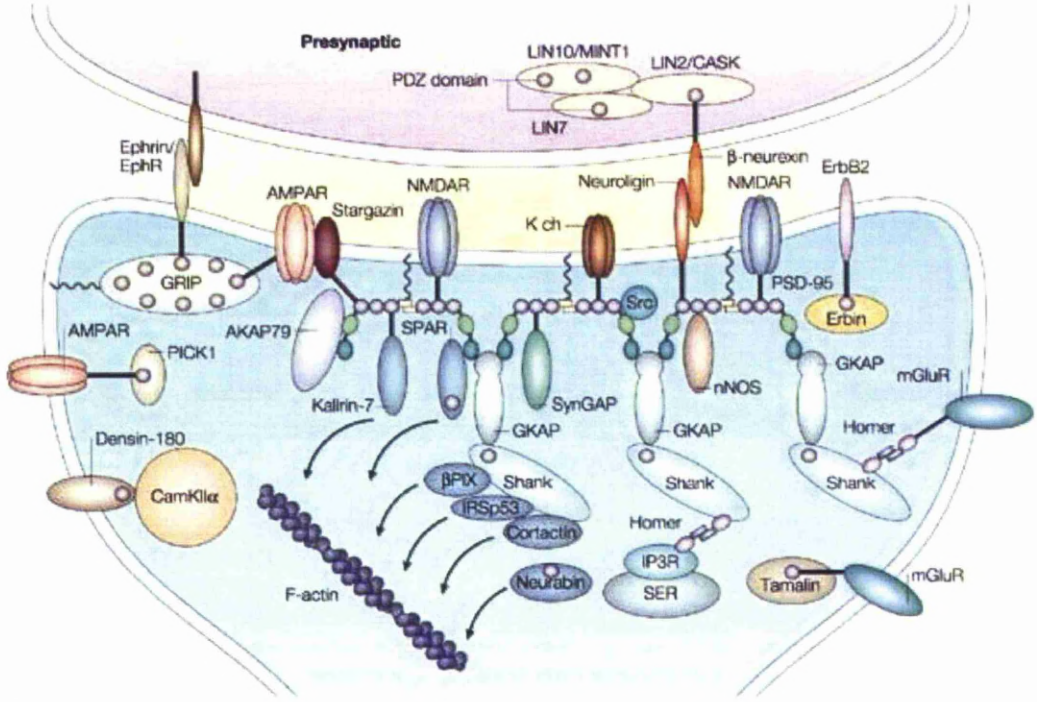


Figure 1.0. A diagrammatic overview of the organization of PDZ proteins at a mammalian excitatory synapse. The schematic of known interactions of PDZ containing proteins at the glutamatergic synapse. The PDZ domains are indicated by purple circles within the scaffold proteins. As this is an overview of a known mechanism each individual protein in the diagram will not be discussed in detail (Kim and Sheng 2004).

Scaffold proteins often contain more than one binding domain. Commonly found domains are the src homology 2 (SH2) or src homology 3 (SH3) domains that can bind either short phosphopeptide motifs or proline rich peptide sequences, in the form PxxP, respectively. A further example is the PDZ domain which is discussed in more detail below (Pawson and Scott, 1997). These multiple binding domains gives the scaffold protein the ability to bind more than one partner at once, helping to assemble the biochemical pathway or protein network.

1.1 MAGUK Proteins

The Membrane Associated Guanylate Kinases (MAGUKs) are a major family of scaffold proteins. They are characterised by a catalytically inactive guanylate kinase like domain (GK) juxtaposed to a src homology 3 (SH3) domain and one, three or five PDZ domains. MAGUKs have been recognised in several species including mammals, drosophila and *C. elegans*. They are located primarily at the post synaptic density (PSD) and tight junctions in neurons although they can be found throughout many organs in the body. Major mammalian members include PSD-95/SAP90, PSD-93/chapsyn-110, SAP97 and SAP102, ZO-1–3, CASK, PALS-1-3, p55-1-7, Carma-1-3 and Magi-1-3 and Caβ-subunit-1-4. Two of these are atypical MAGUKs, Magi substitute the SH3 domain for a WW domain and the Caβ-subunit-1-4 contains an SH3-GK domain but without any PDZ domains and is located at the cytoplasmic C-terminal of a voltage-dependent calcium channel (Dobrosotskaya et al., 1997; Funke et al., 2005; Kornau et al., 1997; Opatowsky et al., 2004; Willott et al., 1993). Figure 1.1 is a schematic representation of some MAGUK proteins, highlighting the domain compositions.

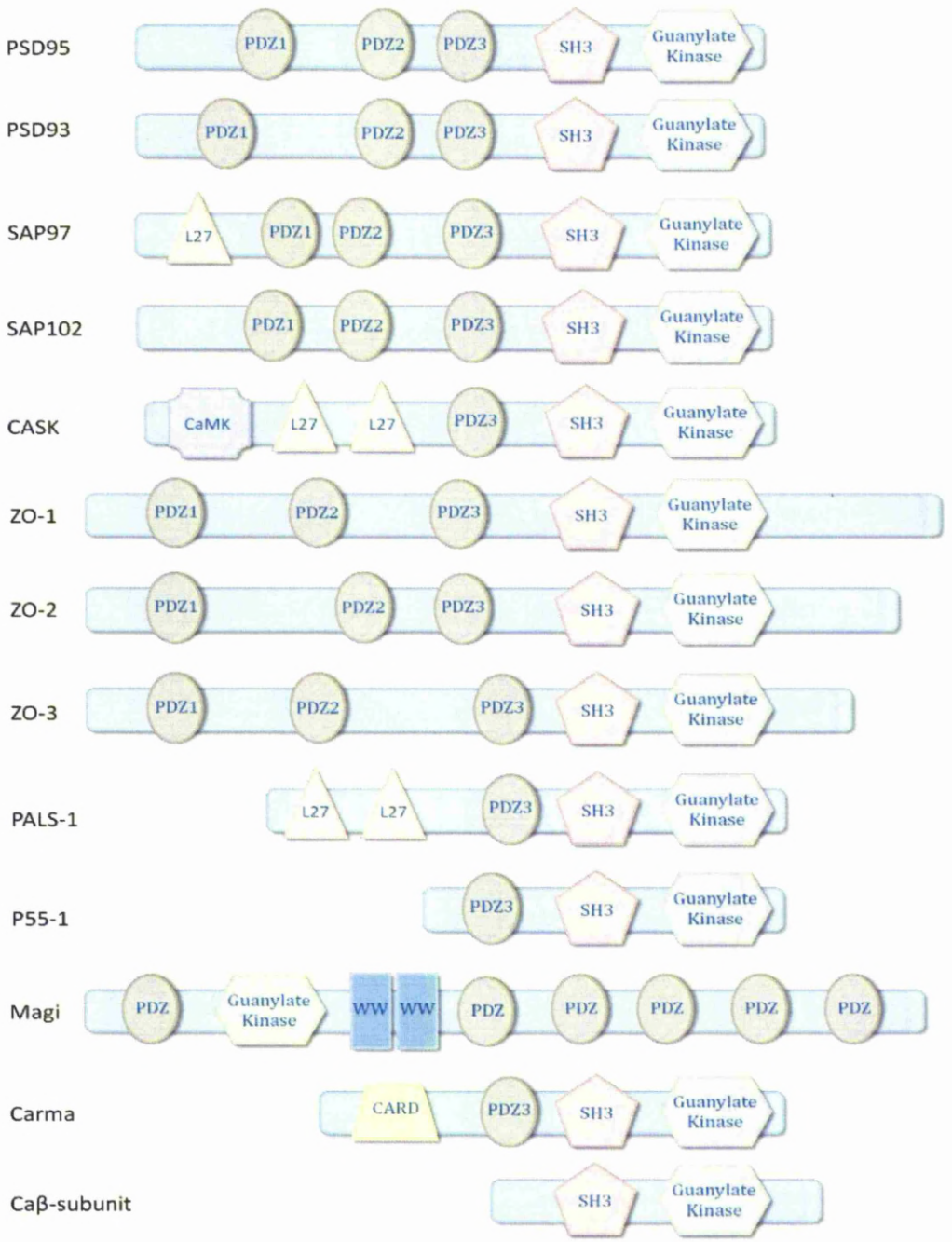


Figure 1.1. Schematic of mammalien MAGUK structures. Common MAGUKs and the domain configuration of each protein. The more unusual domains are L27, a mulitmerisation domain (discussed below), CaMK, calcium-calmodulin kinase, CARD, caspase recuitment domain and WW a double tryptophan motif. The Caβ subunit of the voltage gated ion channel, CaV is also included though not technically a MAGUK protein (Adapted from Funke et al 2005)

1.1.1 Roles of MAGUKs

The plethora of binding domains found in the MAGUKs makes them multifunctional. At epithelial tight junctions, ZO-1 acts dynamically to assemble and regulate the cell boundaries. This is achieved by the simultaneous anchoring of the cytoskeleton to the tight junction, binding the filamentous actin directly and indirectly through binding to α-catenin and cingulin. ZO-1 to 3 are also anchored to the tight junction membrane, the PDZ domains interact with the claudins and the GK domain to occludens, both trans membrane adhesion proteins creating a large complex (Shen et al., 2008). MAGUKs are also instrumental in the clustering of ion channels and receptors together at the postsynaptic density. PSD95, has been shown to cluster the ion channel Kv1.4 (Kim and Sheng, 1996) and the major glutamic recptors, NMDA (Kornau et al., 1997), AMPA 2006 (Beique et al., 2006) and Kainate (Garner et al., 2000).

MAGUKs are also important in signal transduction pathways. Once PSD95 has clustered the NMDA receptor the secondary messenger, nNOS (Ca^{2+}/calmodulin activated nitrogen oxide synthases) is localised at the NMDA receptor through interactions with the second PDZ domain of PSD95 (Brenman et al., 1996). With all the proteins in place localised increases in Ca^{2+} may result in localised increases in nitric oxide that will open the NMDA receptor.

Much of the previous work on MAGUKs has focused on PSD95; in this work investigations will be carried out for the lesser studied close family member, SAP97.

1.2. MAGUK Structure and Functions

To understand MAGUKs function firstly it is necessary to understand the function of all the constituent parts before looking at the whole. Outlined below will be the most common individual MAGUK domains but will, were possible concentrate on those from SAP97.

1.2.1 PDZ Domains

The PDZ domain is one of the most abundant proteins in the genome estimated to be found in over 250 different proteins and is involved in protein – protein interactions (Hung and Sheng, 2002). The term PDZ was derived from the first letters of the three original proteins in which it was discovered, PSD-95, Drosophila large Disc protein and protein ZO-1, all MAGUK proteins (Kennedy, 1995). Containing approximately 90 amino acids they were characterised by a Gly-Leu-Gly-Phe (GLGF) repeat found within its sequence (Cho et al., 1992), which when it was first solved (Cabral et al., 1996) was found to play a major role at the binding site, as a cradle within the binding loop, Figure 1.2 (Doyle et al., 1996).

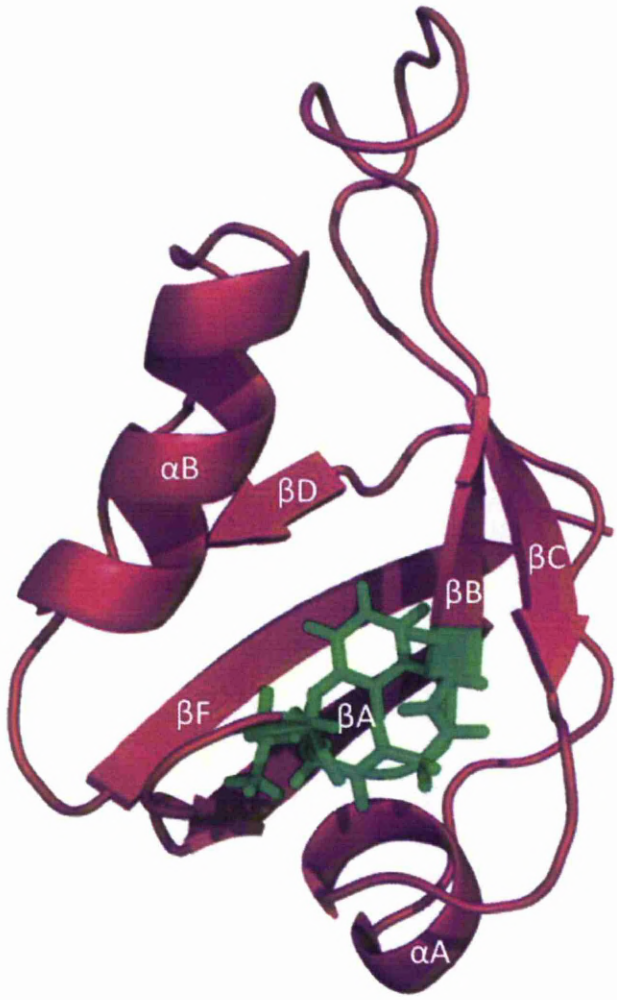


Figure 1.2. Ribbon structure of PDZ 1 domain from SAP97. The GLGF motif is shown at the base of the binding cleft (in green). Adapted from Wang et al 2005.

The structure of the PDZ domain is well conserved in mammalian cells with 6 β-sheets (A – F) and 2 α-helices (A and B). Each PDZ domain contains a specific binding site or pocket formed by βB and αB with the αB running anti parallel to the βB sheet, figure 1.2. The loop that follows the βB-sheet contains the GLGF motif that is essential in stabilising the binding interaction (Doyle et al., 1996).

The binding pocket was shown to have specificity for four C-terminal residues from the binding partner protein, with the basis of the specificity determined by the first residue in the βB sheet and the second residue of the terminal peptide that forms the basis of the PDZ classification. Three general classes of PDZ domain have now been discovered; Class I domains, which recognize the carboxyl terminal motif S/T-X - Φ, where X is any amino acid and Φ is a hydrophobic amino acid, Class II domains, which recognize the carboxyl-terminal motif X-Φ -X- Φ, and Class III domains, which recognize X-D/E-X- Φ as their preferred carboxyl-terminal motif (Hung and Sheng, 2002). Only Class I domains are found in SAP97.

1.2.2 SH3 – GK domains

Along with the PDZ domain the SH3-GK domains are present in all MAGUKs. Figure 1.3 shows a sequence alignment of various SH3-GK domain. The Src homology 3 or SH3 domain is a small protein – protein interaction domain that binds sequences rich in prolines favouring the PxxP core motif where x can be any amino acid (Li, 2005). However, in MAGUKs the SH3 domain rarely binds to proline rich motifs (Funke et al., 2005). The GK domain is homologous to the yeast guanylate kinase domain that catalyses GMP to GDP with ATP (Berger et al., 1989) but in MAGUKs it is catalytically inactive possibly due to missing three amino acids in the ATP binding site (Lue et al., 1994).

The first structure of the SH3-GK domain was elucidated in PSD95 (McGee et al., 2001; Tavares et al., 2001). The structure showed two unexpected features of the SH3-GK domain. Firstly, that the SH3-GK domains were intimately linked, with

the SH3 domain being a split domain, its final β-sheet formed from the C-terminus polypeptide chain found beyond the GK domain (Figure 1.4). Secondly, the Hook region which connects the SH3 and GK domains forms part of the SH3 domain, with the Hook α-helix masking the usual site of protein – protein interaction on the SH3 binding domain. The latter feature explains the lack of interactions to proline-rich motifs (McGee et al., 2001).

Recently a crystal structure of the SH3-GK domains of ZO-1 was solved, Figure 1.5. Initially they used the full domains but the crystals diffracted poorly so they removed the U5 Hook region that was poorly resolved in PSD95. They were then able to solve the structure and without the use of molecular replacement using PSD95. The resulting structure gave unbiased domain positions (Lye et al., 2010). The refined structure showed the SH3 and GK domains as separate units that were joined by a spilt domain as in PSD95. However, the two domains were more open in conformation than in PSD95. This was attributed to a cis-proline (P_{714}) in PSD95 that lies between the split domain β–sheet and the last GK α-helix (αVI) and which is replaced by a glutamine in ZO-1 (Lye et al., 2010). Interestingly, the cis – proline is present in SAP97 suggesting a more closed conformation akin to PSD95.

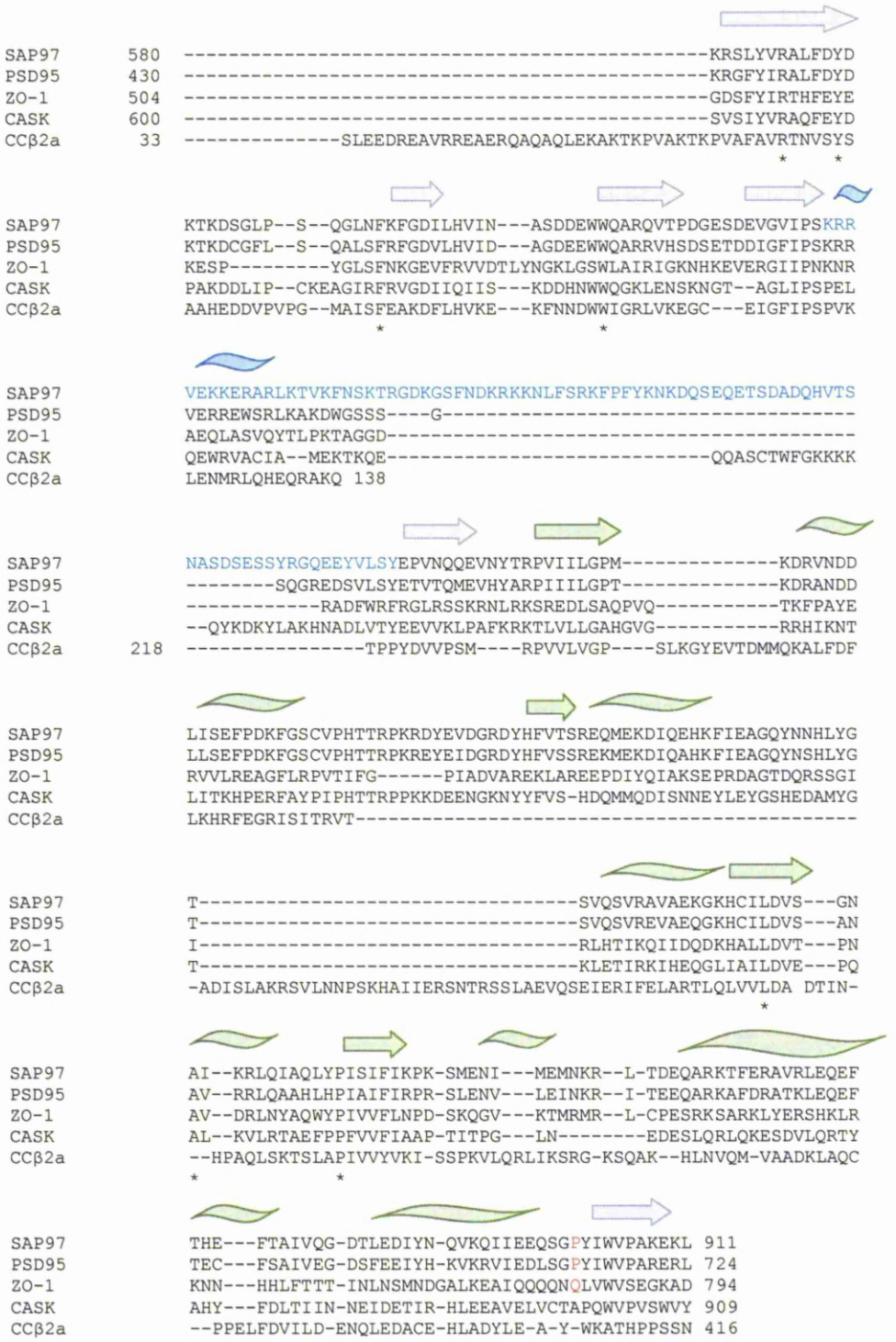


Figure 1.3: see next page for full description.

Figure 1.3. Sequence alignment of the SH3, Hook and GK domains. Alignment of different proteins in the SWISS-PROT database using ClustalW2. Arrows and spirals depict areas of β-sheet and α-helix of PSD95 (McGee et al., 2001). The GK domain is in green, SH3 magenta and Hook cyan, note the hook sequence is also cyan. * indicates conserved residues and red letters show PSD95 P714 that is suspected to bring the SH3 and GK domains closer together (Lye et al., 2010).

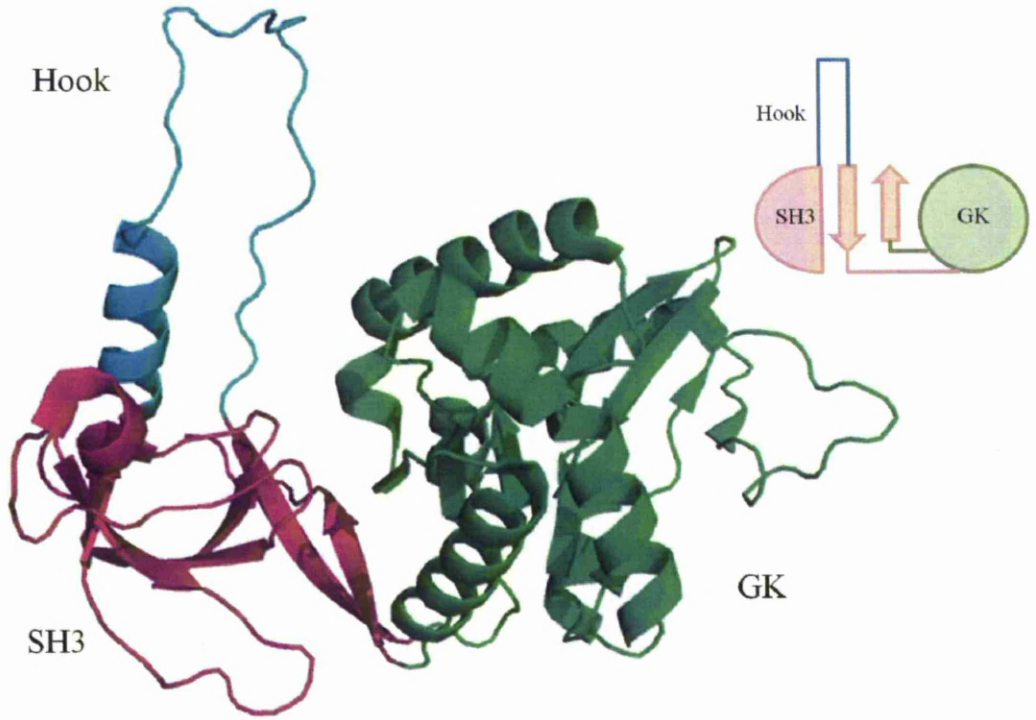


Figure 1.4. PSD95 SH3-GK Domains. The SH3 (magenta) and GK domains (green) of PSD-95 share a β-sheet and the Hook (cyan) region contains an α-helix that separates structural components of the SH3 fold. The insert is a schematic representation of the SH3-GK module showing how the Hook divides the SH3 domain into two sub-domains. Adapted from (McGee et al., 2001).

In the ZO-1 structure the Hook region was truncated but using the poorly diffracted SH3-GK domain it was possible to show that the truncation did not affect the structural conformation. An unpublished structure of ZO-3 (3KFV) with intact Hook domain was deposited into the protein data bank (PDB). This also showed an open conformation about the spilt domain similar to ZO-1 and when overlaid with the PSD95 and ZO-1 it showed that the inter-domain angle of the Hook helix was variable through each conformation suggesting that the Hook regions variable

length and angle in various different MAGUKS is related to its function (Lye et al., 2010).

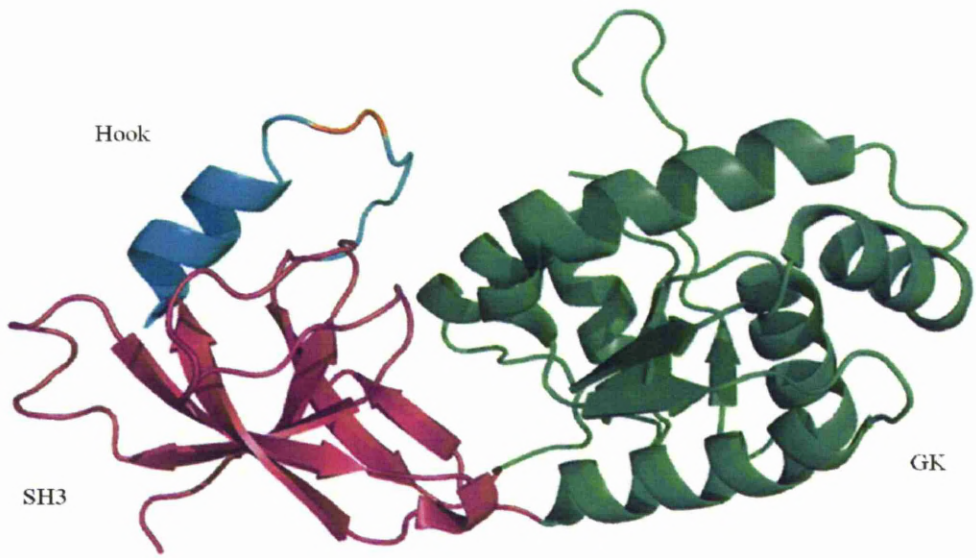


Figure 1.5. ZO-1 SH3-GK Domain. The SH3 (magenta) shows an open conformation in relation to the GK domain (green). The Hook (cyan) region is truncated, orange indicates the site of truncation.

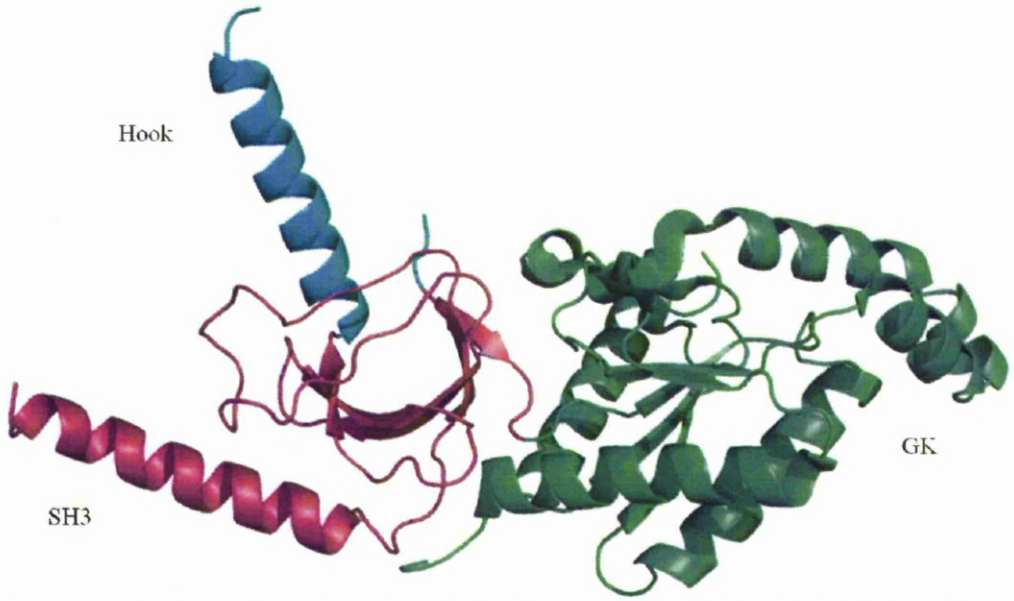


Figure 1.6. Calcium voltage channel, CaV β-subunit-2a SH3-GK domain. The SH3 (magenta) shows a long –helix at the amnio terminal end followed by part of the Hook (cyan) region and then GK domain (green).

Another crystal structure of an SH3-GK domain has been solved, that of voltage gated ion channel, CaV β-subunit-2a (Opatowsky et al., 2004). The structure showed an extra, long α-helix at the amino terminus, with smaller β-sheets and longer loop regions in the SH3 domain than when compared to PSD95 and ZO-1. The GK core domain was similar to PSD95 but its ability to bind small nucleotides was compromised. Unlike PSD95 and ZO-1 there was no split domain between the SH3 and GK domains and as a result the SH3-GK domains showed an even more open conformation and with the Hook region directed away from the GK domain, Figure 1.6.

Table 1.0; MAGUK SH3-GK domain binding partners and locations

SH3-GK Domain binding	Binding partner	Location	Reference
SAP97	AKAP79	SH3	(Colledge et al., 2000)
	Calmodulin, CaM	Hook	(Paarmann et al., 2002)
	GAKIN	GK	(Hanada et al., 2000)
	GKAP	GK	(Wu et al., 2000)
	Protein 4.1	Hook	(Lue et al., 1994)
PSD95	AKAP79	SH3	(Colledge et al., 2000)
	GAKIN	GK	(Hanada et al., 2000)
	GKAP	GK	(Wu et al., 2000)
	kainate	SH3 and GK	(Garcia et al., 1998)
ZO-1	α -Catenin	SH3 and Hook	(Muller et al., 2005)
	Calmodulin, CaM	GK	(Paarmann et al., 2002)
	Occludin	GK and Hook	(Schmidt et al., 2004)
	ZONAB	SH3	(Lye et al., 2010)
CaV β-subunit-2a	AID-CaV α -subunit	GK	(Hidalgo and Neely, 2007)
	Anhak	SH3/GK	(Buraei and Yang, 2010)
	Ca^{2+}activated K^{+}channel	GK	(Buraei and Yang 2010)
	Dynamin	SH3	(Hidalgo and Neely 2007)
	RIM1	SH3 Hook GK	(Buraei and Yang 2010)
	Zinc Transporter 1 (ZnT-1)	Unknown	(Buraei and Yang 2010)

Several binding partners for the SH3-GK domains mentioned above have been identified and are shown in Table 1.0. Some of the large binding proteins such as occludin and α-catenin were shown to bind across two domains. While CaM which has been shown to bind to the Hook domain of SAP97, interacts with the GK domain in ZO-1. The SH3-GK domain of CaV β-subunit-2a interacts with the α-

interaction domain (AID) from CaV, an α-helical domain that binds tightly into a groove on GK. More recently interactions with several other binding partners have been discovered. These interactions are predominantly other ion channels including the Large-conductance Ca^{2+}-activated K^{+} channel (BKCa), Zinc Transporter 1 (ZnT-1) and Bestrophin, a retinal chloride channel. Binding of other proteins such as the large scaffolding protein, Anhak and the GTPase, dynamin have also been reported (Hidalgo and Neely 2007; Buraei and Yang 2010).

1.2.3 L27 Domain

The L27 domain was first identified in the C. *elegans* Lin-2 and Lin-7 proteins before mammalian orthologs were found in MAGUK proteins SAP97, CASK and PALS (Doerks et al., 2000). Studies of the L27 domain of SAP97 showed a propensity to form homomultimers (Nakagawa et al., 2004). This was followed by NMR structures of L27 domain of SAP97 and the L27 domain of CASK that showed the formation of a heterotetramer, Figure 1.7 and identified the mechanism of formation; the individual L27 domains were found to exist in an unfolded state, only forming α-helices upon the formation of a heterodimer and subsequently the heterotetramer. What was also noted was that the α-helices of both SAP97 and CASK were very similar in structure even though the sequence similarity was only 16% (Feng et al., 2004). The mulitmerisation is thought to increase the clustering of SAP97 and CASK at the plasma membrane.

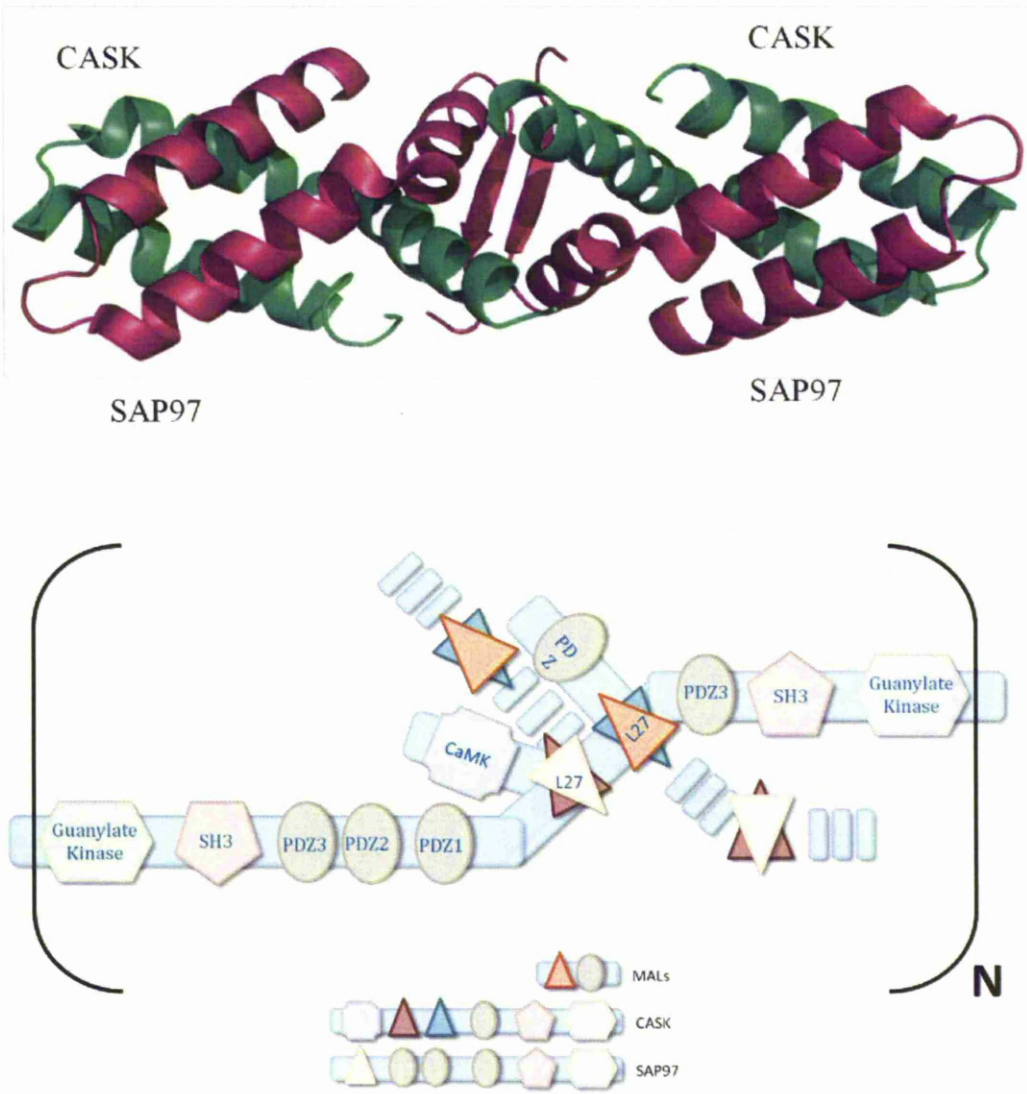


Figure 1.7. L27 hetero-tetrameric complex in SAP97 and CASK. Top, shows the two SAP97 L27 domains (magenta) interacting with the two CASK L27 domains (green). Bottom shows a proposed model for the polymerisation of SAP97 with CASK and a third protein MALS. The SAP97 L27 binds with the first of the CASK L27 domains while the second CASK L27 domain binds to MALS forming a heterodimeric unit which then goes to form hetero-tetrameric interaction (shown by the dashed lines). Adapted from Feng et al 2004.

1.3 SAP97

A principal member of the MAGUK family, synapse associated protein, SAP97 is localised to the cytoskeleton and expressed in many cell types including the heart and kidneys, but is found extensively at the synapses where it was first identified (Lue et al., 1994). SAP97 contains an L27, three PDZ and SH3-GK domains, all joined by unique unstructured linker domains named U1 – U5 that are thought to provide flexibility within the protein, Figure 1.8 (Wu et al., 2000). The U5 / Hook linker region has also been shown to contain several mRNA splice variants leading to different isoforms of SAP97 that are suggested to play a role in tissue specificity with the U5 I4 isoform found only in brain and liver (McLaughlin et al., 2002). SAP97 is also known to have two sites of phosphorylation at Ser-29 in the L27 domain and Ser-232 in the PDZ1 domain that are both known to be phosphorylated by calcium/calmodulin kinaseII (CaMKII) that is suggested to play a role in trafficking the SAP97 to the plasma membrane (Mauceri et al., 2007).

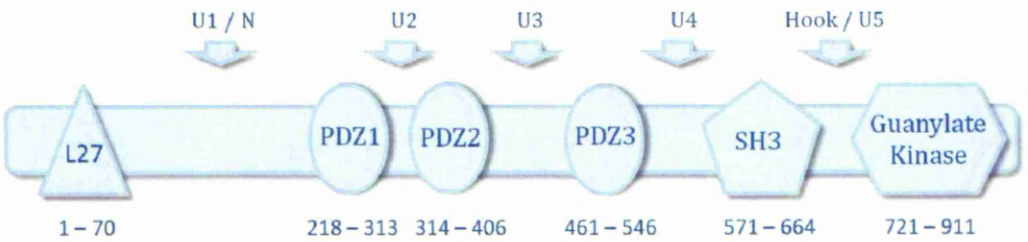


Figure 1.8. Schematic of SAP97 showing the locations of each domain. Individual domains are joined together by linker regions termed U regions that are unstructured.

1.3.1 SAP97 Structure

There are no high resolution structures of intact full-length SAP97. Individually however, the structure for all three PDZ domains of SAP97, have previously been characterised, PDZ1 by NMR (Wang et al., 2005) and PDZ2 and PDZ3 by crystallography (Cabral et al., 1996; von Ossowski et al., 2006; Zhang et al., 2007). The structure of SH3-GUK domain remains unsolved in SAP97 although the structure of the SH3-GK from the related MAGUK protein PSD95 have been solved, see above (McGee et al., 2001; Tavares et al., 2001).

In 2000, Wu et al. reported a computational model of SAP97 using the known crystal structures of similar PDZ domains and molecular models of all the subunits. They used the modelled structure to explain the various interactions between SAP97 and target proteins binding. The model that was generated showed SAP97 as a compact structure with intramolecular interactions between domains. Furthermore, they demonstrated that these interactions were regulated, that is the domains associate and dissociate depending on what SAP97 interacts with, Figure 1.9. The model suggested interactions between the SH3 and GK domains, later shown to occur in PSD95 (McGee et al., 2001) and interactions between Hook and GK domains such that the Hook domain would sterically hinder the binding of GKAP to the GK domain and that this could be regulated by the binding of the L27 or N (U1) domain to the Hook (Wu et al., 2000).

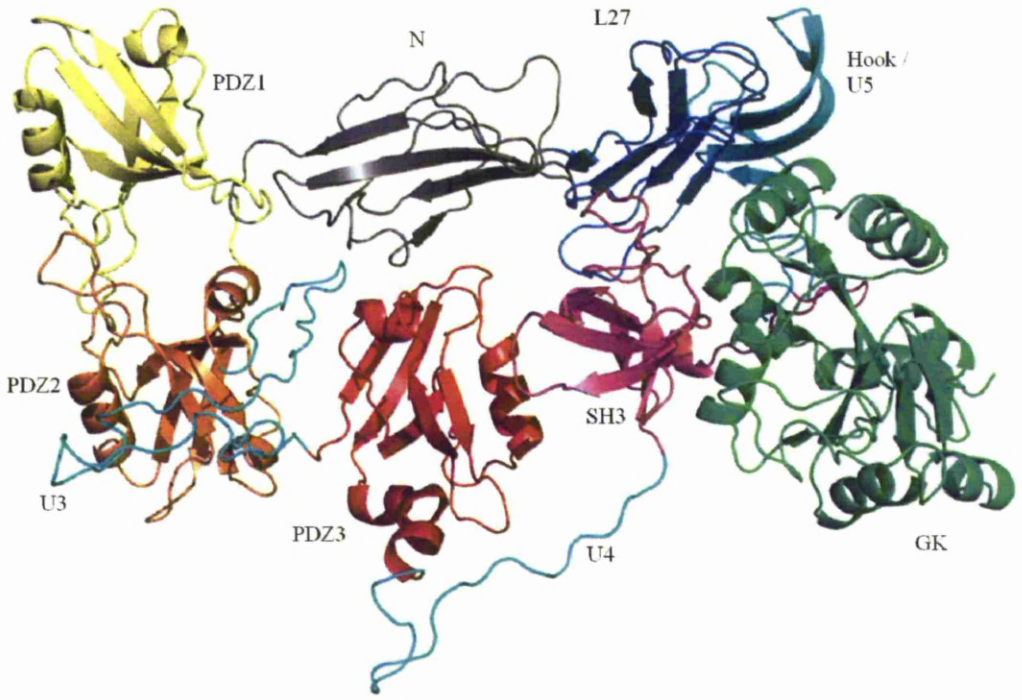


Figure 1.9. Computational model of SAP97. Model derived using the program FTDOCK with the individual domains taken form SAP97 or domain homologues found by blast. GK in green, SH3 in magenta, PDZ3 red, PDZ2 orange, PDZ1 yellow, U3,4,5 in cyan, N grey and L27 in blue. Adapted from Wu et al 2000.

Although there is some support for a compact SAP97 as predicted by Wu et al 2000, the low resolution structure electron micrographs (EM) showed only 35% of the particles formed a compact structure whereas for the majority 65%, elongated rods were observed (Nakagawa et al., 2004). The same group also showed that SAP97 usually occurred in a dimeric form mediated through the L27 domain (Nakagawa et al., 2004).

Conformation dynamics within SAP97 was determined more recently for the PDZ region. Two constructs of PDZ12 and PDZ123 were analysed by small angle X-ray scattering and solution NMR, Figure 1.10. The data showed that there was little

interdomain conformational movement in PDZ12, most likely due to the short U2 linker region. Similar observations were also made in the NMR studies of PSD95 PDZ12 (Long et al., 2003). Greater conformational movement and freedom was observed in the PDZ123. This was expected due to the presence of the longer U3 linker region (Goult et al., 2007).

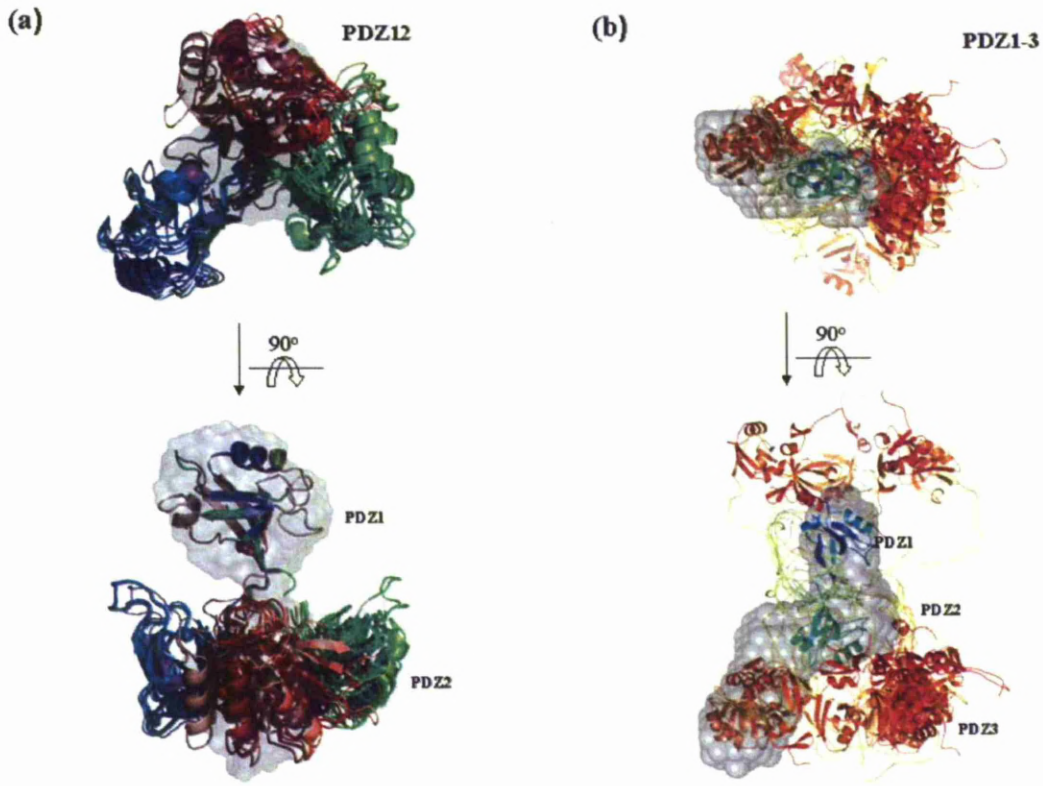


Figure 1.10. Structural conformation variation of PDZ1-3 domain from SAXS data. 20 rigid body SAXS models (BUNCH) are superimposed over a low resolution ab initio surface envelope (grey). A, shows PDZ12 overlaid and aligned on PDZ1. B shows PDZ123 aligned over PDZ12 (blue/green) then superimposed over the surface envelope, showing that PDZ3 (red) is highly mobile, taken from Goult et al. 2007

1.3.2 SAP97 Interactions

With several distinct binding domains and located in several different tissues it is of no surprise that SAP97 binds to many different partners. Consequently, there are many suggested roles for SAP97 within the cell. Outlined below are a few these.

1.3.2.1 PDZ-mediated interactions and functions

One of the first roles discovered for the PDZ domains was the clustering and trafficking of potassium channels (Craven and Bredt, 1998; He et al., 2006), specifically the potassium inward rectifier, Kir2.1 (Leonoudakis et al., 2001).

Kir2.1 is a tetrameric membrane spanning ion channel with a large cytoplasmic pore domain that contains several sites of phosphorylation by PKA and PKC. In addition, PtdIns(4,5)P_2 and Mg^{2+} binds to and blocks the channel. The C-terminus has a PDZ-binding motif ESEI (Bichet et al., 2003). Each of the SAP97 PDZ domains were shown to bind to the C-terminal sequence ^{418}EPRPLRRESEI^{428} of Kir2.1, although it was shown that PDZ2 had the greatest affinity (Goult et al., 2007). SAP97 was also shown to cluster the Kv1.4 potassium channel at the cell membrane through its PDZ domains (Kim and Sheng, 1996). The PDZ domains were also indicated to cluster the glutamate receptor AMPA at the plasma membrane (Leonard et al., 1998).

1.3.2.2 SH3-GK-mediated interactions and functions

The SH3 domain that was shown to bind the scaffolding anchor protein, AKAP79 (see below) albeit in a proline independent manner (Colledge et al., 2000), whilst

the Hook linker region between the SH3 and GK domains binds to calmodulin, CaM (Paarmann et al., 2002). The binding of either AKAP79 or CaM may regulate the scaffolding action of SAP97; CaM has been suggested to sterically hinder the binding of protein 4.1, a cytoskeletal binding protein that also binds to the Hook region (Colledge et al., 2000; Paarmann et al., 2002). Whether CaM and AKAP79 can bind SAP97 together or whether they oppose each other has not been investigated. More recently, the SH3 domain of SAP97 has also been found to bind to α-secretase (Marcello et al., 2007).

The SAP97 and PSD95 GK domains have both been shown to interact with guanylate kinase-associated protein, GKAP. No function for the interaction with SAP97 has been found so far although GKAP has been shown to colocalise with PSD95 to cluster Kv1.4 ion channels (Kim and Sheng, 1996).

SAP97 has also been shown to localise to the cytoskeleton in epithelial cells initially though binding of the Hook region to Protein 4.1 (Lue et al., 1994). However, it was then shown that interactions with the cytoskeleton occurred primarily through the L27 domain and not through associations with protein 4.1 (Wu et al., 1998).

SAP97 has been studied in many different tissues, each identifying different binding partners and functions. This shows that SAP97 is a versatile MAGUK protein able to perform different roles, forming complexes and interactions with different proteins when required throughout the cells when required.

1.4 A-Kinase Anchoring Protein, AKAP

A second major family of proteins involved in scaffolding at the plasma membrane are the A-Kinase Anchoring Proteins (AKAPs). These belong to a diverse family of more than 50 proteins that were originally differentiated by their molecular weight such as AKAP18, 79, 220 and 350 but also contain the proteins gravin and WAVE1 (Wiskott–Aldrich verprolin-homology protein-1). They are characterised by a protein kinase-A (PKA) anchor domain and are able to compartmentalise to specific sites in the cell through interactions with other scaffolds that enables PKA to be distributed and bind to or interact with other signalling or scaffold molecules (Welling, 2008; Wong and Scott, 2004).

1.4.1 Roles of AKAP

Many ion channels are regulated by phosphorylation from broad acting kinases and phosphatases and the AKAPs play a major role here as a scaffold protein. In the inactive state, the regulatory (R) subunits of PKA bind to the C-terminal region of AKAP. Upon activation by cAMP, PKA is released in its catalytically active form from AKAP. Active PKA proceeds to phosphorylate other proteins in its vicinity (Carr et al., 1992; Uhler et al., 1986).

More details of the precise mechanism of the process described above has only been revealed recently (Kinderman et al., 2006). cAMP-dependent Protein kinase A

contains two subunits, one catalytic (C) and the second regulatory (R). The N-terminal of the R subunit contains a small binding domain commonly called the dimerisation/docking domain (D/D). This is composed of an anti-parallel four α-helix bundle that provides a docking surface for AKAP. The R subunit binds AKAP through a 14 – 18 amino acid amphipathic helix (an α-helix with opposing hydrophobic and hydrophilic faces), Figure 1.11 (Carr et al., 1992; Kinderman et al., 2006). At C-terminus of PKA, two tandem cAMP binding domains are present; when cAMP binds to these sites, a conformational change occurs in the R subunit which releases the C subunit (Kinderman et al., 2006).

The R subunit is also found in two isoforms RI and RII which confer specificity for different AKAPs, though the D/D domain fold is conserved in each subunit (Gold et al., 2006).

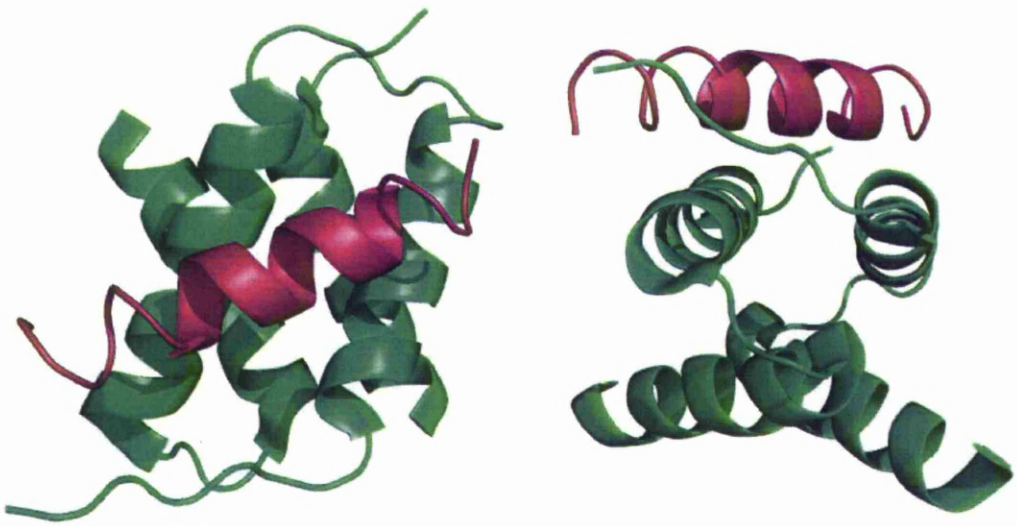


Figure 1.11. RIIα subunit of PKA dimerisation/docking domain in complex with D-AKAP2. Structures 90° to each other showing peptide D-AKAP2$_{639-649}$ bound to the D/D domain of the RIIα subunit of PKA. Adapted from Kinderman et al 2006.

The compartmentalisation of AKAPs is vital to the specificity of the PKA catalysis with AKAPs located at the plasma membrane, in nuclei, mitochondria, microtubules and the actin cytoskeleton. Two mechanisms have so far been identified. The first tethers AKAPs to membranes via interactions with phospholipids (AKAP79), myristoyl groups (gravin) or by both myristoyl and palmitoyl groups (AKAP18). The second uses splice variants that direct different protein-protein interactions; for example, AKAP350 binds to the NMDA receptor but the AKAP450 elongated splice variant binds to the centrosome (Wong and Scott, 2004).

AKAPs ability to bind to other signalling molecules and enzymes allows them to orchestrate signal transduction. This multiple binding was first established using yeast two hybrid assay in 1995 when it was found that AKAP79 bound PKA and

the Ca^{2+}/calmodulin-dependent protein phosphatase 2B, calcineurin (CaN) simultaneously (Coghlan et al., 1995). A year later PKC was found to bind independently to both PKA and CaN (Klauck et al., 1996).

Analysis of AKAP18 showed that it contained a domain that resembled 2H phosphoesterases (2H from two conserved histidines), which are enzymes that catalyse small nucleotides. Structural characterisation the $AKAP18_{76\text{-}292}$ domain revealed that it interacts with 5'AMP and cytosine monophosphate (CMP) although no phosphoesterase activity was identified, suggesting that it may catalyse another nucleotide or that it acts as an AMP sensor, Figure 1.12 (Gold et al., 2008).

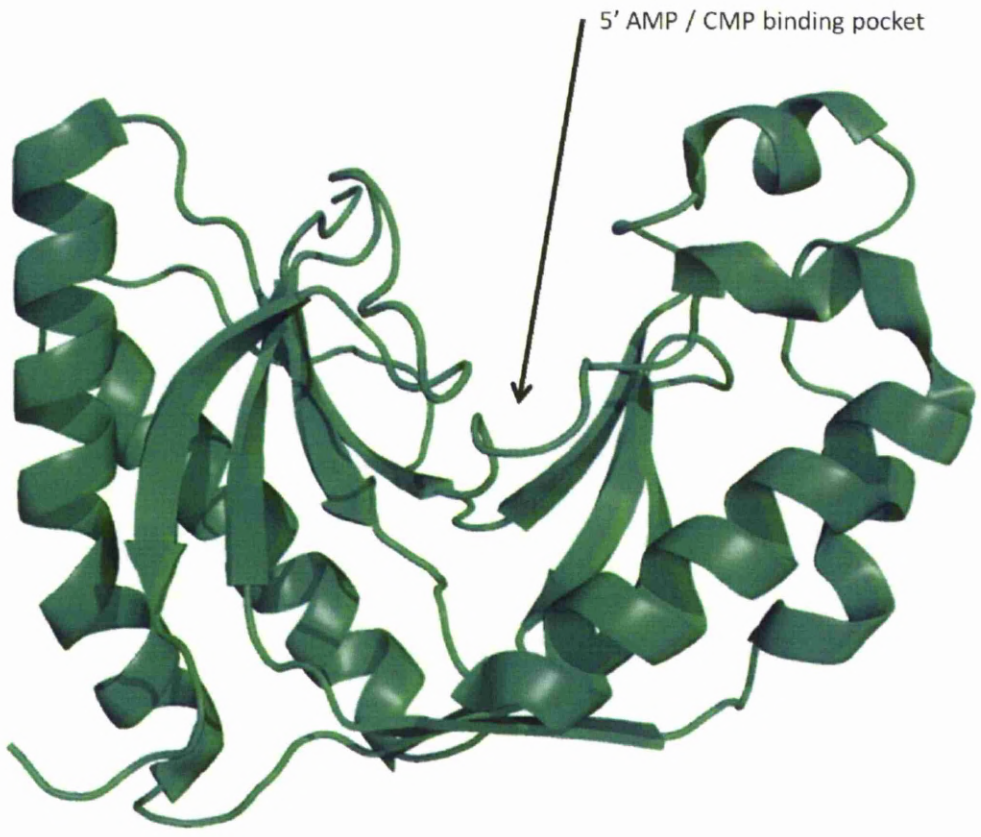


Figure 1.12. AKAP18 central domain. The domain resembles 2H phosphoesterase and binds both 5'AMP and cytosine monophosphate (CMP) in the central binding pocket. No catalysis has been observed. Adapted from Gold et al 2008.

My research will focus on AKAP79. Although one of the most widely studied AKAPs its structural data and binding mechanisms have yet to be explored in full.

1.5 AKAP79

The 427 amino acid protein, AKAP79 (also named AKAP5) has a molecular weight of ~47KD and is a member of a small closely related family of AKAPs along with the murine (AKAP150) and bovine (AKAP75) analogues.

1.5.1 AKAP79 Structure and Function

AKAP79 compartmentalises to the plasma membrane and has been found in various cell types including the post-synaptic density (PSD) of the neuron (Carr et al., 1992), the kidney (Welling, 2008) and heart cells (Dart and Leyland, 2001). The mechanism by which AKAP79 attaches to the plasma membrane is through three N-terminal highly basic regions (named A, B and C) that have been shown to localise to the plasma membrane through fluorescence microscopy and bind to the acidic phospholipid, phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P_2). However, it is not known whether binding to acidic phospholipids is purely electrostatic or specific for PtdIns(4,5)P_2, Figure 1.13 (Dell'Acqua et al., 1998). The N-terminal region was also shown to bind to both PtdIns(4,5)P_2 and F-actin at the dendritic spine (Gomez et al., 2002) and to cadherins in epithelial cell tight junctions (Gorski et al., 2005).

The highly basic A region, $AKAP79_{32-51}$ was shown to bind PKC in an inactive bound form (Klauck et al., 1996). This was interesting on two counts. First, the region was not a known PKC binding sequence, thus it was suggested that this was an unique class of PKC that binds in a phosphotidylserine dependent manner (Faux and Scott, 1997). Second, the A region was already known to be a CaM binding site to AKAP. It was found that CaM/Ca^{2+}compete with PKC for the binding site, providing a mechanism for secondary messenger (Ca^{2+}) regulation for releasing inactive PKC from AKAP79 (Faux and Scott, 1997).

This was further complicated by the fact that AKAP79 binding to membranes and the actin cytoskeleton was shown to be regulated by the binding of CaM or PKC (Dell'Acqua et al., 1998; Gomez et al., 2002; Gorski et al., 2005). More recent work has shown that only AKAP79 compartmentalisation was required for neuronal development at dendritic spines and not the action of PKA or CaN. This was achieved by monitoring the clustering of AMPA receptors to the membrane and actin cytoskeleton using mutant AKAP79 constructs without the PKA or CaN domains (Robertson et al., 2009).

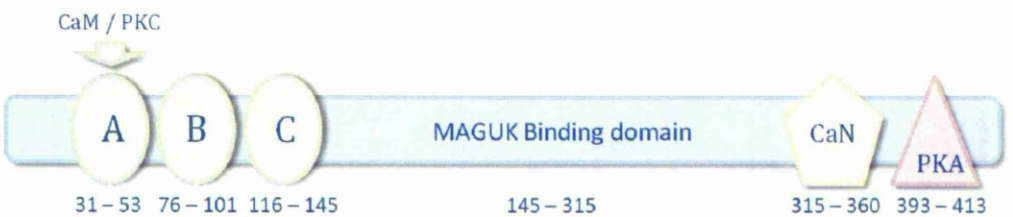


Figure 1.13. Schematic of AKAP79. Showing the locations of binding sites. A, B and C represent highly basic regions. Exact MAGUK binding domain is unknown but is suggested to lie within region shown.

The binding complex of AKAP79 to the RII subunit of PKA has been solved by solution NMR (Newlon et al., 2001). The group expressed the RII binding domain (1-44) of the R subunit dimer and synthesised the amphipathic helix from $AKAP79_{393-413}$. They found that the amphipathic helix bound at 45° to the dimer in a hydrophobic pocket by the same method as D-AKAP2 shown previously (Figure 1.14).

Figure 1.14. RIIα subunit of PKA dimerisation/docking domain in complex with AKAP79. Structures 90° to each other showing peptide $AKAP79_{392-413}$ (magenta) bound to the D/D domain of the RIIα subunit of PKA (green). Adapted from Newlon et al 2001

No further structural characterisation has been carried out on AKAP79 (or any other AKAPs). This is most likely due to a lack of secondary structure in the central region and difficulties in expressing AKAPs in bacteria (Gold et al., 2008)

1.5.2 AKAP79 Interactions

AKAP79 has been shown to associate and bind with several partners including PKA, PKC and CaN (Klauck et al., 1996). More recently binding with other proteins has been identified including CaM (Dell'Acqua et al., 1998), SAP97, PSD95 (Colledge et al., 2000) and Kir2.1 (Dart and Leyland, 2001). However, through its ability to form simultaneous interactions with more than one binding partner AKAP has a scaffolding role in creating protein networks much akin to SAP97.

Through fluorescence resonance energy transfer (FRET) microscopy analysis and patch clamping it was shown that AKAP79 was able to assemble a network of, PKA, CaN and the L-type Ca^{2+} channel, $Ca_V1.2$ all localised to the cell membrane. The experiments also showed that the PKA and CaN, respectively, phosphorylated and dephosphorylated the $Ca_V1.2$, that is, directly opposing each other (Oliveria et al., 2007). Further research also using FRET showed that a complex of AKAP79 and SAP97 was able to localise to the cell membrane with the PDZ domains of SAP97 associated to the β-adrenergic receptors (β_1-AR). This enabled SAP97 to facilitate in the PKA mediated phosphorylation of Ser_{312} on the β-adrenergic receptor (Gardner et al., 2007).

Complexes of AKAP79 and SAP97 have also been linked to localising and assembling complexes with glutamate receptor-1 domain (GluR1) of AMPA

(Colledge et al., 2000). Further to this work it has been found that AKAP79 also enhances the regulation of PKC phosphorylation of GluR1 (Tavalin, 2008).

As mentioned previously, SAP97 has been shown to associate with the K^+ Channel Kir2.1 (Goult et al., 2007). Interestingly, AKAP79 has also been shown to bind directly with Kir2.1 thus possibly anchoring PKA closer in proximity to Kir2.1 (Dart and Leyland, 2001). Following on from the AKAP79-SAP97-β_1AR assembly, one could assume that a similar assembly of AKAP79-SAP97-Kir2.1 would also form although this complex remains to be proven either *in vitro* or *in vivo*.

1.6 Calmodulin

Nearly all cellular processes utilize the divalent cation Ca^{2+} as a secondary messenger. Ca^{2+} effects are mediated through the highly conserved, 17 kDa calcium binding protein, calmodulin, CaM (Klee et al., 1980), a protein that has been shown to interact with both SAP97 and AKAP79.

1.6.1 Roles of Calmodulin

CaM is an important regulatory protein that modulates the activity of signalling molecules through its Ca^{2+} sensitivity and has been shown to interact with over one hundred different proteins (Xia and Storm, 2005). The diversity of physiological roles is also large ranging from the comprehensively studied CaM dependent serine

and threonine proteases to the anthrax adenylate cyclase. The diversity is attributed to the abundance and conservation of the protein across eukaryote species and even to the prokaryote, *bacillus anthracis* (Hoeflich and Ikura, 2002).

1.6.2 Calmodulin Structure

The CaM structure contains N and C-terminal lobes joined by a flexible linker region. On each lobe there are two EF hand motifs that allow CaM to bind 4 Ca^{2+} molecules, Figure 1.15. Saturation with Ca^{2+} leads to a conformational change that facilitates the binding to many enzymes and proteins transducing a signal to the target (Swulius and Waxham, 2008).

The solved structure of CaM was first reported in 1988 at 2.2 Å in the calcium bound form (Babu et al., 1988). This was followed by a 1 Å model by Wilson 2000 that shows the flexible central helix open for binding (Wilson and Brunger, 2000).

The crystal structures also showed that the EF hands were folded into a helix-loop-helix motif and that the loop. The Ca^{2+} ions are coordinated to the EF hands by the 12 amino acids loop through side chain oxygen atoms from amino acids in positions 1, 3, 5 and 7. The coordination sphere is completed by a H_2O molecule, Figure 1.15 (Grabarek, 2006).

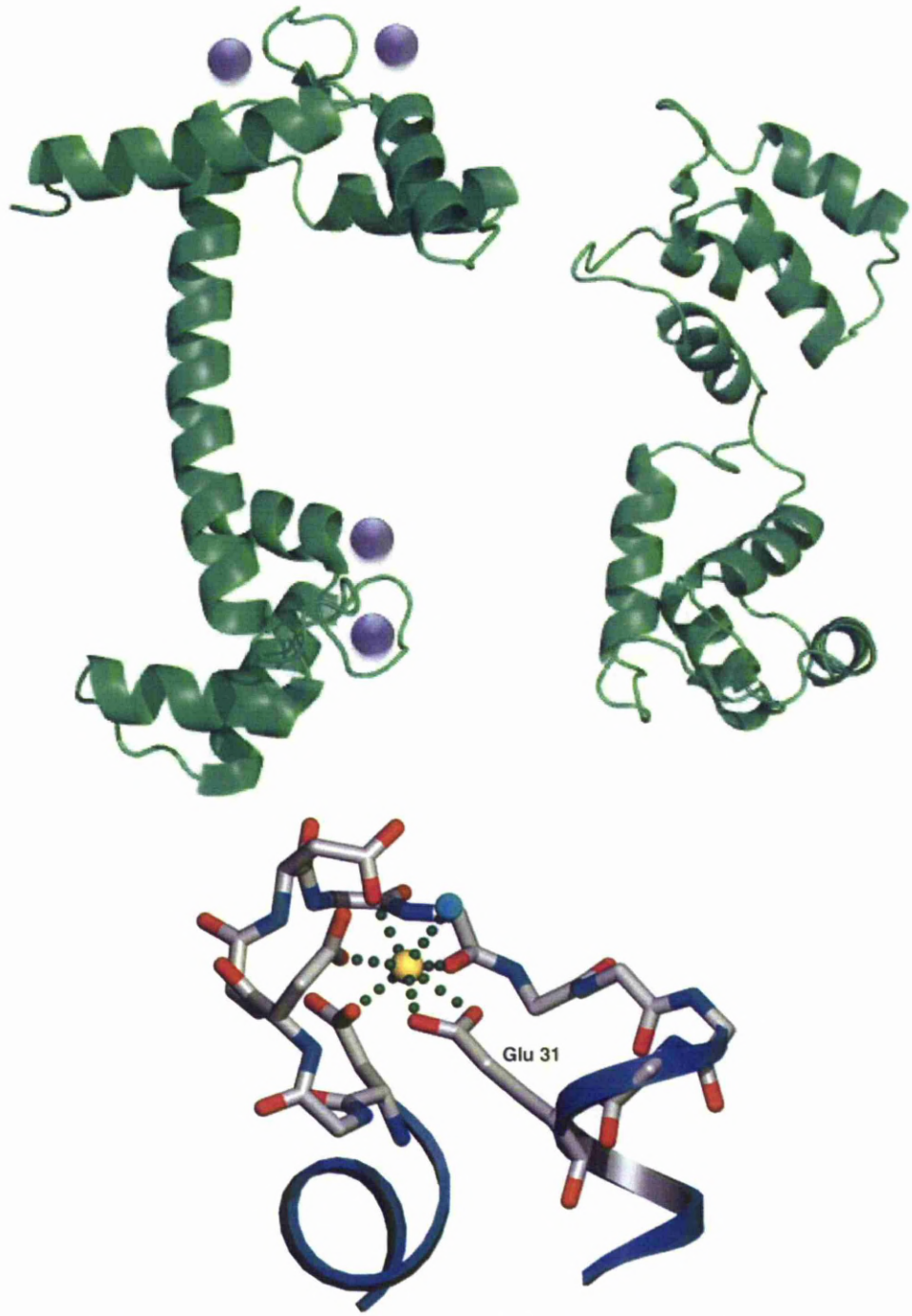


Figure 1.15. A ribbon representation of CaM, Left shows CaM (from Paramecium *tetraurelia*) bound with Ca^{2+} represented as magenta spheres derived from 1Å X-ray crystallography, showing the extended dumbell conformation when bound to Ca^{2+} (Wilson et al 2000). Right shows the apo CaM (from Xenopus *laevis*) solved by NMR in the closed state (Zhang et al 1995). Bottom shows the helix-loop-heilx motif from an EF hand showing the coordination of the Ca^{2+} cation, taken from Garbarek 2006.

When CaM is without any Ca^{2+} ions it is said to be in the apo state whose structure is very different from the Ca^{2+}-bound state (Figure 1.15). The angles of the α-helices of the EF-hands motifs increase by 36° - 44° which closes access to the central region (Zhang et al., 1995).

The CaM model can be thought of as being like a pair of head phones; when Ca^{2+} is bound, the ear pieces are pulled apart allowing you to place said 'headphones' on. However, when the Ca^{2+} is absent the 'headphones' close shut.

1.6.3 Calmodulin Function

Calmodulin contains a high proportion of methionine residues, 9 from 148. These methionines have been shown to form a malleable nonpolar binding surface, this coupled to long flexible side chains provides exposed hydrophobic patches that are involved in targeted binding (Vogel et al., 2011).

The binding surfaces of CaM have been shown to bind to a basic amphipathic helix that contains positive residues flanked by and possibly interspersed with hydrophobic residues but the target binding sequences often showed no homology. However, a pattern between target sequences was discovered. Several binding motifs were found in which the hydrophobic side of the α-helix interacted with the hydrophobic residues on CaM. Motifs 1-8-14 and 1-5-10 were first identified, were the numbers relate to the position of the hydrophobic residue in the target sequence, Figure 1.16 (Meador et al., 1993; Rhoads and Friedberg, 1997; Vogel et al., 2011). More recently 1-12 and 1-16 motifs have been identified but these are rarer. There

was also a calcium independent motif discovered; the IQ that contains the motif IQxxxRGxxxR. Overall in 2000 there were approximately 180 different sequences shown to bind CaM (Yap et al., 2000)

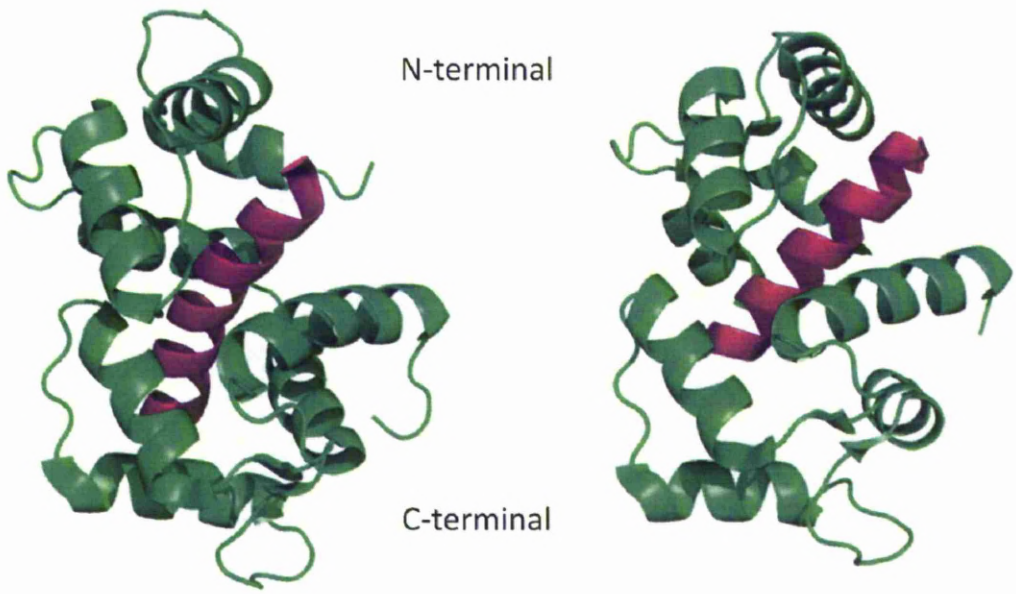


Figure 1.16. Ribbon representation of CaM Complexes, Left shows CaM comlpexed with CaM kinase I peptide in a 1-14 binding mode, adapted from pdb 2L7L (Vogel et al., 2011). Right shows CaM comlpexed with CaM kinase II peptide in a 1-5-10 binding mode, adapted from pdb 1CDM (Meador, Means et al. 1993).

CaM has been shown to bind to both SAP97 and AKAP79 (Faux and Scott, 1997; Paarmann et al., 2002). In SAP97, CaM binds to the Hook region that may interfere with the binding of both the cytoskeletal protein 4.1 and AKAP79, that may regulate SAP97s clustering ability (Colledge et al., 2000; Paarmann et al., 2002). In AKAP79, an influx in secondary messenger Ca^{2+} activates CaM which competes with PKC to bind to the highly basic region, regulating the release of inactive PKC from AKAP79. The Ca^{2+} also has the effect of releasing the dephosphatase,

calcineurin. CaN which binds to AKAP79 (Faux and Scott, 1997; Oliveria et al., 2007).

1.7 Summary

SAP97 and AKAP79 are two scaffolding proteins that interact together and with other regulatory proteins including CaM, PKA, PKC and with the receptor and ion channel Kir2.1.

Kir2.1 is a member of a large family of potassium ion channels. The potassium inward rectifier, is predominantly expressed in the heart where maintains the resting membrane potential, buffers external K^+, and repolarises cardiac muscle action potential through its strong inward rectifier function (Doupnik et al., 1995; Kubo et al., 2005).

The structure contains two main constituents, a tetrameric membrane spanning pore region and a large intracellular cytoplasmic pore region formed by the carboxy- and amino-terminal domains. As mentioned previously the cytoplasmic domains contains several sites of phosphorylation by PKA and PKC. In addition, PtdIns(4,5)P_2 and Mg^{2+} binds to and blocks the channel. The C-terminus has a PDZ-binding motif ESEI that has been shown to interact with SAP97 (Bichet et al., 2003).

There are no published mechanisms for a SAP97-AKAP79-Kir2.1 complex, although, they have been shown to interact with each other. Recent physiological

work on Kir channels *in vivo* using patch clamping did suggest that PKC inhibits Kir while PKA activate the channels (Park et al., 2008). However, there was no mention of the mechanisms by which these regulations were achieved, it could be possible that AKAP79 and CaM are involved. Below is a theoretical model of the SAP97-AKAP79-Kir2.1 complex, Figure 1.17.

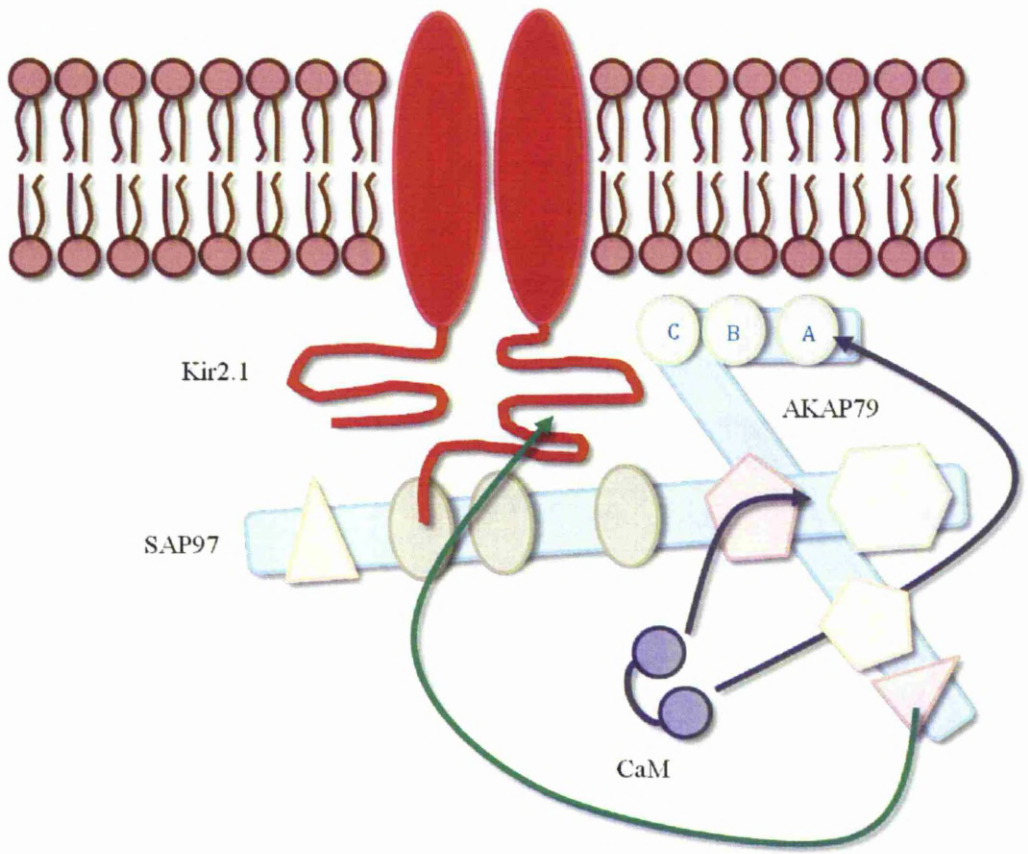


Figure 1.17. Possible model for SAP97-AKAP79-Kir2.1 complex. The Kir2.1 (red) is clustered to the plasma membrane by SAP97 PDZ domains. AKAP79 is then able to anchor at the phospholipids and interact with the SAP97 that could enable PKA to phosphorylate Kir2.1 (green arrow). An increase in CaM may regulate the complex by releasing the AKAP79 from the phospholipids and possibly interfering with the SAP97 AKAP79 interactions (purple arrows).

1.8 Aims of Project

The overall goals of this thesis are to obtain comprehensive structural information on (i) AKAP79 and (ii) SAP97 and to understand how both these proteins are regulated through their interactions with other proteins. To achieve these goals, the specific aims are as follows.

The Biophysical Characterisation of AKAP79. This will be done by making constructs for *E.coli* expression of full length and separate domains of AKAP79, protein purification and characterisation using NMR, SECMALLS and circular dichroism were relevant. This will be followed by the examination of protein-protein interactions in the SAP97-AKAP79 system. This includes AKAP79 interaction with SAP97, determination of binding affinities using Isothermal Calorimetry (ITC) and by identifying residues important for interactions using NMR chemical shift mapping.

Investigating the interactions of AKAP79 with Calmodulin. This will identify residues important for interactions using NMR chemical shift mapping and by determining binding affinities using Isothermal Calorimetry (ITC)

The Biophysical Characterisation of SAP97. This will be done by expressing, purifying and characterising the different domains of SAP97 including Δ70, Δ461, Δ546, GK and SAPN1 domain using NMR and / or SECMALLS This will be followed by the examination of the protein-protein interactions between SAP97 and Calmodulin by determining binding affinities using fluorescence spectrophotometry and by identifying residues important for interactions using NMR chemical shift mapping. The protein-protein interactions between SAP97 and Kir2.1 C-terminus fragment will then be examined by expressing and purifying $^{13}C/^{15}N$ C368, and interpreting the assigned spectra when complexed with SAP97 PDZ2.

Finally SAP97 will be characterised using small angle X-ray scattering (SAXS) to investigate and interpret the large multi domain SAP97 Δ70, Δ461, Δ546 constructs.

CHAPTER 2

TECHNICAL INTRODUCTION TO SMALL ANGLE X-RAY SCATTERING

2.0 Introduction

Small angle X-ray scattering (SAXS) is a powerful method that allows the study of biological macromolecules in solution. SAXS can be used to analyse peptides to viruses to protein complexes in physiological or denaturing conditions, without the need to crystallise. This technical introduction will outline: (a) the basic theory of SAXS and (b) the analysis strategy of the scattering. Most important the different computer programmes that are commonly used for data analyses are described in some detail.

2.1 Scattering

The basic principal of SAXS is to pass a collimated (parallel), monochromatic beam of X-rays through a solution of highly purified, monodisperse macromolecules. The intensities of the scattered X-rays are measured as a function of the scattering angle (Figure 2.1). The wave vector of the X-ray beam is $K = 2\pi/\lambda$ where λ is the X-ray wavelength. The isotropic scattered intensity, I, is recorded as a function of the momentum transfer, $s = 4\pi\sin\theta / \lambda$ where 2θ is the angle between the incident and scattered beam. The scattering from the background buffer solution is measured separately and subtracted from the protein solution (Svergun, 2010)

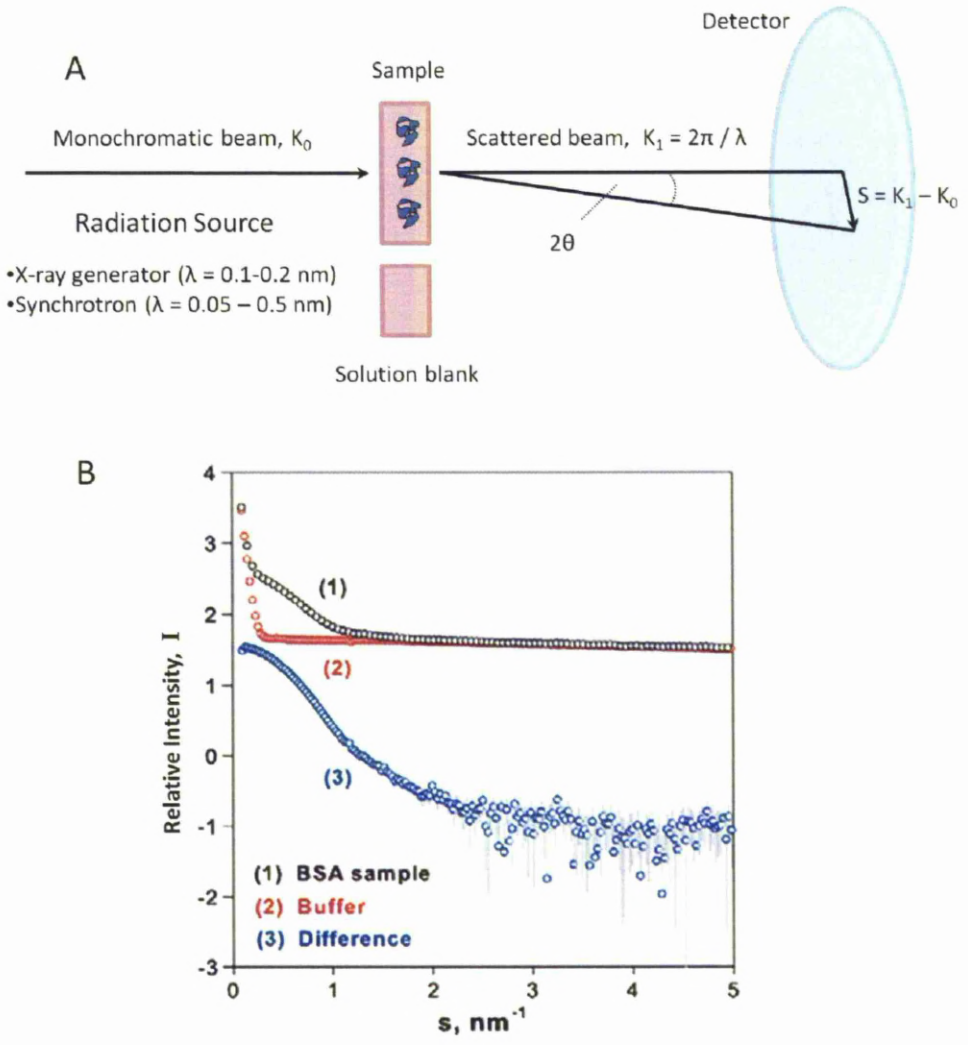


Figure 2.1. Schematic of basic SAXS experiment. A, shows a simplified standard experimental set up. B, shows the resultant scattering data for a BSA standard. Adapted from (Mertens and Svergun, 2010).

2.2 GNOM

The initial scattering data can be analysed for several parameters by the computer program, GNOM (Svergun, 1992). These include the regularization parameter, P(r) that gives information on the shape of the proteins in solution such as whether it is spherical or rod shaped. It represents the histogram of distances between pairs of points within protein and is related to the maximum dimension, D_{max} of the protein (Koch et al 2003). P(r) is measured indirectly because any small errors in the scattering would cause large errors in P(r); this is compounded by the finite number of intensities that are measured during scattering. To increase the accuracy further the program GNOM uses *a priori* assumptions about the protein in question which include the D_{max} of the protein which can then be refined later to improve the quality of the solution.

Finally GNOM will also calculate an accurate value of the radius of gyration, Rg of the protein, a measure of the size of the protein derived from the root mean square distance of the central particle to the scattering length density distribution (Koch et al., 2003; Svergun, 1992).

2.3 Shape determination

2.3.1 GASBOR

Recent modelling approaches have allowed the construction of low resolution three dimensional models directly from scattering data through *ab initio* modelling. Initially, a single envelope model (SASHA) was created. This was followed by a bead model (DAMMIN) but this too has been superseded by a dummy residue model, GASBOR. This method uses the data generated by GNOM and the number of amino acids in the protein to create the model and tries to use fewer parameters than other modelling programs to increase calculation speed. Each amino acid is represented by a dummy residue (DR) and during the iterative process each DR is randomly moved to an arbitrary point, 0.38 nm from the next DR within a shell of constraints provided by D_{max} and the scattering data until a solution is generated (Figure 2.2). This method does not provide a unique solution but a manifold of configurations that correspond to nearly the same scattering pattern. The differences between each model are less than the previous methods (Svergun et al., 2001). The models from each individual run can be superimposed or averaged using another program, DAMAVER that measures the normalised spatial discrepancy (NSD) for each residue in each run and then compares each run to each other, removing outlying runs and superimposing and averaging the other runs to give a final average 3D model

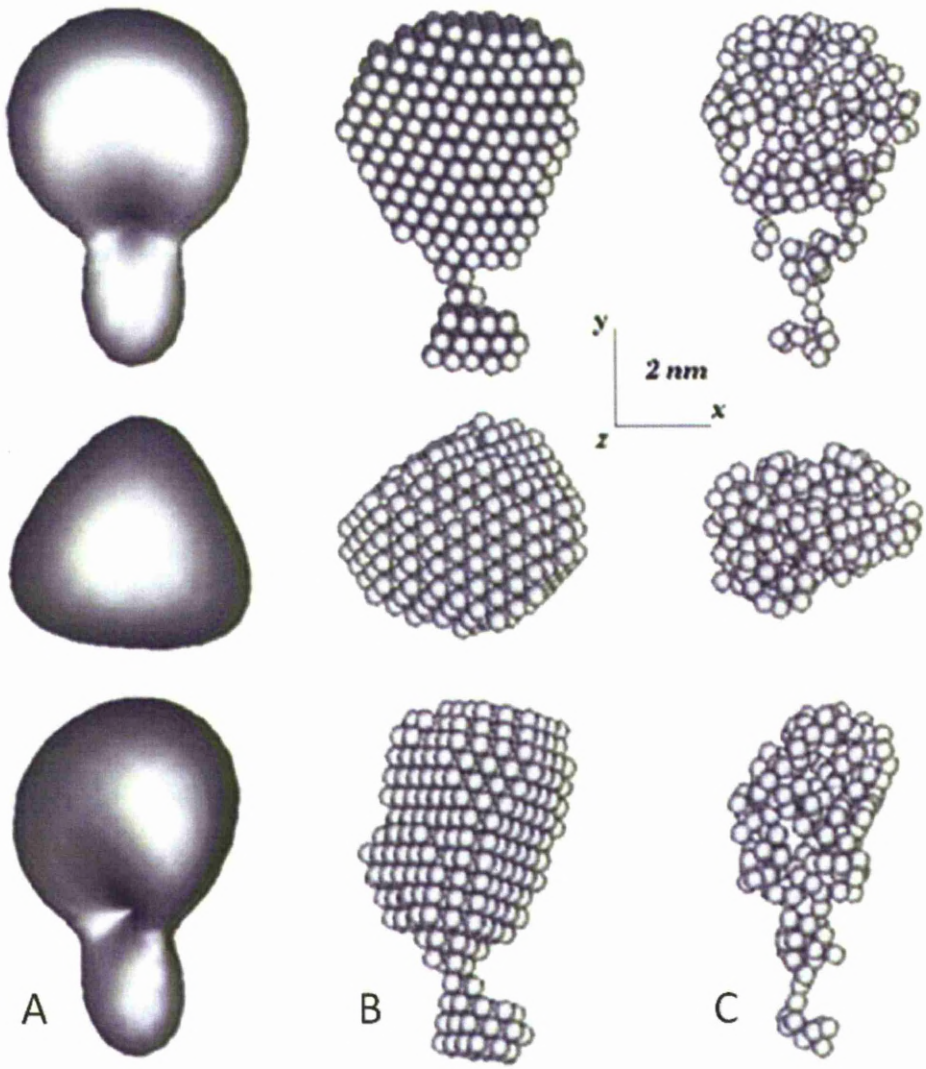


Figure 2.2. Ab initio models of Chitin binding protein, CHB1 used to show differences between different 3D modelling programs. A, uses an envelope model from SASHA, B uses the bead model, DAMMIN and C uses the dummy residue method, GASBOR. Models are rotated through 90°. Taken and adapted from Koch, Vachette et al. 2003.

2.4 Rigid Body Modelling

Large macromolecules often comprise of many individual domains. To study these macromolecules, the individual domains could be fitted directly into the low resolution *ab initio* GASBOR models. However, it is more reliable to use a rigid body modeller program such as BUNCH or EOM that assemble models using direct refinement of the scattering data (Mertens and Svergun, 2010).

2.4.1 BUNCH

Bunch uses scattering data to build models of multi-domain proteins or protein complexes from X-ray crystal or NMR structures, or homology models of the subunits and domains (Petoukhov and Svergun, 2005). The models allow for proteins or domains to be connected without steric clashes and can incorporate missing linker regions. The method uses heuristic algorithms, that is an algorithm of best fit or best solution for the model is optimised iteratively to find the best solution amongst all possible solutions. It does not however guarantee the best solution is found and is often considered an approximate solution. Added to this the linker regions processed by *ab initio* modelling using dummy residues (DR) are often assumed to be rigid (Mertens and Svergun, 2010).

2.4.2 Ensemble Optimisation Method, EOM

This method has been developed to analyse flexible systems such as unfolded proteins or multi domain proteins with long flexible linkers (Bernado et al., 2007). Instead of using the usual method of searching for a single model or conformation that fits the experimental scattering data and which adheres to physical and biochemical scrutiny, EOM was developed to look at flexible proteins that may exist in a mixture of forms.

The EOM consist of two parts. Firstly an object is represented by an ensemble of many different conformers. The ensemble is usually 10000 conformers that are randomly generated but possible conformations to cover the conformational space. The average scattering of the ensemble is then found by averaging each scattering profile of each individual conformer. Secondly, a genetic algorithm is then employed for a pool of subsets and further evolved for 5000 generations to complete the ensemble optimisation. The structures were generated from the ensemble and were then analysed to quantify their diversity (frequency) and distribution dependent on the radius of gyration (R_g) of the ensemble.

2.5 Homology Modelling

The rigid body modellers require high resolution crystal structures for their refinements. Where these are not available, homology models are created using the high resolution structures of related proteins domains or folds as templates. Two

main types of homology modelling exist, *ab initio* and comparative and these are described in detail below.

2.5.1 Ab initio Modelling – I TASSER

I-TASSER is a computational model run through a web server that uses a combination of techniques including *ab initio* modelling to predict 3D structures of proteins from just their amino acid sequence and then predict their function (Roy et al., 2010). The program methodology progresses through 4 stages which are briefly described below:

Stage 1- Threading; initially, bioinformatics techniques, such as BLAST identify evolutionary relatives and then a multiple sequence alignment to predict the secondary structure is defined. This alignment is then threaded through a "meta-threading server" called LOMETS (houses seven individual threading programs) that searches a representative PDB library for similar solved structures or structural motifs to match the secondary structure prediction. The best predicted models are then chosen based on the standard deviation of the energy values relative to all the alignments (Z-score).

Stage 2-Assembly; Fragments of the threading alignment structure template that aligned well are excised and used to assemble a conformational model. To this are added the areas that did not align to any structural motif (loops and tails) generated through *ab initio* modelling. A Monte Carlo simulation technique is used to assembly the fragments guided by PDB statistical terms (hydrophobicity, H-bonds,

Cα correlations) spatial restraints from threading templates and sequence based contact prediction. Conformations are then clustered and the lowest free energy state conformers or "decoys" identified. The 3D coordinates of all the clustered decoys are then averaged to form a "cluster centroid", essentially a cluster of many different conformations with the lowest free energy that are averaged together.

Stage 3-Model selection and refinement; the cluster centroids are sent back to the stage 1 fragment assembly for a second iteration, this time pooling external restraints from the threading alignments on the decoys in the cluster. This enables the removal of steric clashes and refines the topology. The decoys are then clustered again and those with the lowest energy state are inputted into a refinement program, REMO, that generates the final structure using an all atom model that is built on Cα and H-bonding networks.

Stages 4- Structural based functional annotation; to attempt to elucidate the function of the queried protein the newly defined structure is matched to other proteins of known structure and function in the PDB.

The five best predicted structures are then presented, their qualities assessed by their C-scores which are determined by a calculation based on the number of decoys in the cluster compared to the total, the RMSD of the decoys to the cluster average and the Z-score.

2.5.2 Comparative modelling (CM) – MODELLER

This program generates the most probable model structures of a sequence using the alignment to other known structures as templates. The process uses a set of spatial restraints that are replicated from the template to the new model. These include Cα – Cα and main chain N – O distances and main chain and side chain dihedral angles (Sali and Blundell, 1993).

The program itself runs off scripts written in the Python computer language and is run from the command prompt. The modelling process follows a logical progression;

- Sequence inputted
- Template identified / inputted
- Sequences aligned
- Models generated
- Models Evaluated

Models are generated and these are evaluated using the DOPE (**D**iscrete **O**ptimized **P**rotein **E**nergy) score that uses probability theory based on the atomic distance statistical potential of the model and the template (Shen and Sali, 2006).

2.6 CRYSOL

When high resolution protein or domain structures or models generated from these structures, by *ab initio* or computer modelling methods are compared to the original

scattering data the structural similarity needs to be validated. CRYSOL is a program that was developed to evaluate and validate the models to SAXS profiles by using curve fitting to the scattering profiles. The program calculates a scattering pattern using the experimentally derived scattering amplitudes and the hydration shell of the atomic group of the residues (Svergun et al., 1995).

2.7 Computational summary

The methods described above involve many computational steps, these have been outlined in the flow diagram below, Figure 2.3.

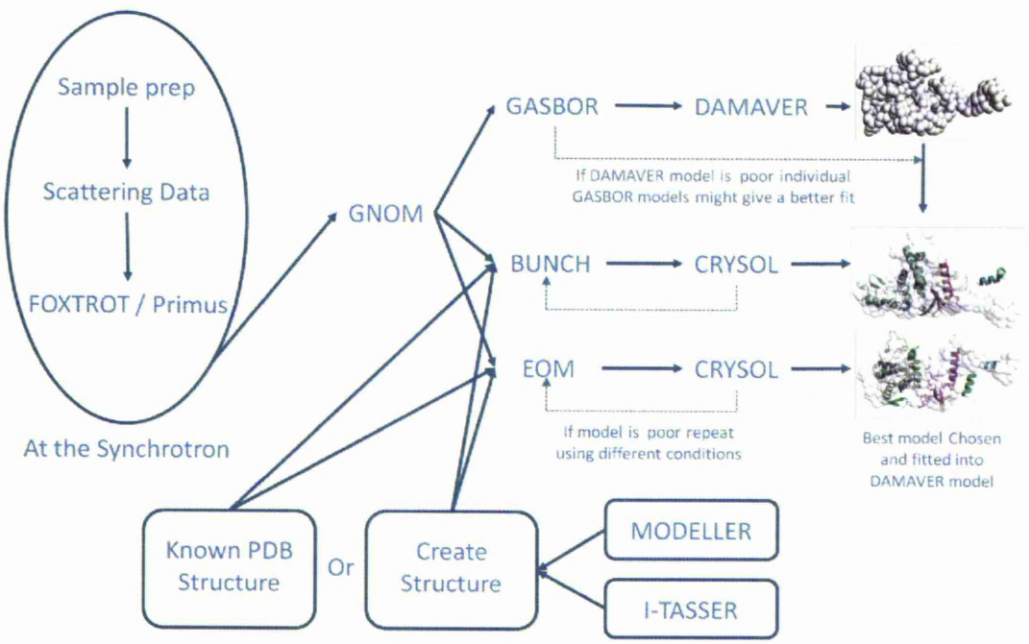


Figure 2.3. Flow diagram of SAXS interpretation. Initial data collection and data processing is carried out at the synchrotron or X-ray source. The remaining interpretation requires freeware programs and can be processed on any modern computer.

CHAPTER 3

MATERIALS AND METHODS

3.0 Introduction

Throughout this study many methods of cloning, protein expression and purification have been used and developed for several different protein constructs of SAP97 and AKAP79. To allow the discussion to follow the method directly, the method development and final methods for the SAP97 and AKAP79 protein constructs will not be in this chapter but will be discussed in full in Chapters 4 (for AKAP9) and 6 (for SAP97).

This chapter will concentrate on other proteins used in the study, assays and other methods used throughout this thesis.

3.1 Chemicals and Reagents

PCR reagents were purchased from Novagen, DNA miniprep kits were from QIAGEN. Luria Broth (LB), yeast extract and agar were from Merck while the tryptone was from Melford and agarose was purchased from Bioline. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was from Melford along with dithiothreitol (DTT), ampicillin (Amp), kanamycin (Kan) and Tris HCl. Chromatography columns and media were purchased from GE Healthcare. ^{15}N Ammonium chloride, ^{13}C Glucose and D_2O were purchased from Goss Scientific. Unless otherwise stated, all other chemicals used were from Sigma-Aldrich and were of the correct purity required.

3.2 Protein Constructs.

T7 vectors were used throughout the study, the AKAP79 constructs were all created as part of this project while all the other constructs were supplied by Dr. Mark Leyland, University of Leicester, Table 3.1 gives a full list of all the DNA constructs used in this thesis.

Table 3.1 Summary of all constructs. Including purification tags, specific cleavage site and vector to be cloned into. TEV = Tobacco Etch Virus protease, SUMO = Small Ubiquitin MOdified protein and PP = Precission Protease

Protein	Construct	Vector
AKAP79 Full Length	HexaHis-TEV-1 – 427	pET M11
AKAP79 N	HexaHis-TEV-1 – 153	pET M11
AKAP79 M	HexaHis-TEV-153 – 315	pET M11
AKAP79 C	HexaHis-TEV-315 – 427	pET M11
AKAP79 M Sumo	SUMO-HexaHis-SUMO cleavage-153 – 315	pOPIN-S
$AKAP79_{19\text{-}61}$	SUMO-HexaHis-SUMO cleavage-19-61	pOPIN-S
$AKAP79_{19\text{-}52}$	SUMO-HexaHis-SUMO cleavage-19-52	pOPIN-S
$AKAP79_{25\text{-}61}$	SUMO-HexaHis-SUMO cleavage-25-61	pOPIN-S
SAP97 Δ70	HexaHis-TEV-70 – 911	pLEIC-01
SAP97 Δ461	HexaHis-TEV-461 – 911	pLEIC-01
SAP97 Δ546	HexaHis-TEV-546 – 911	pLEIC-01
SAP97 GK	HexaHis-TEV-721 – 900	pLEIC-01
SAP97 N PDZ1	HexaHis-TEV-70 – 313	pLEIC-01
SAP97 PDZ2	GST-PP-313-412	pGEX-6P
Kir2.1 C368	GST-PP-368-428	pGEX-6P

3.3 AKAP79 Cloning

The vectors pET-M11 (Pinotsis et al., 2006) and pOPIN-S (Berrow et al., 2007) were used in these studies. The vector maps are shown below, Figure 3.1.

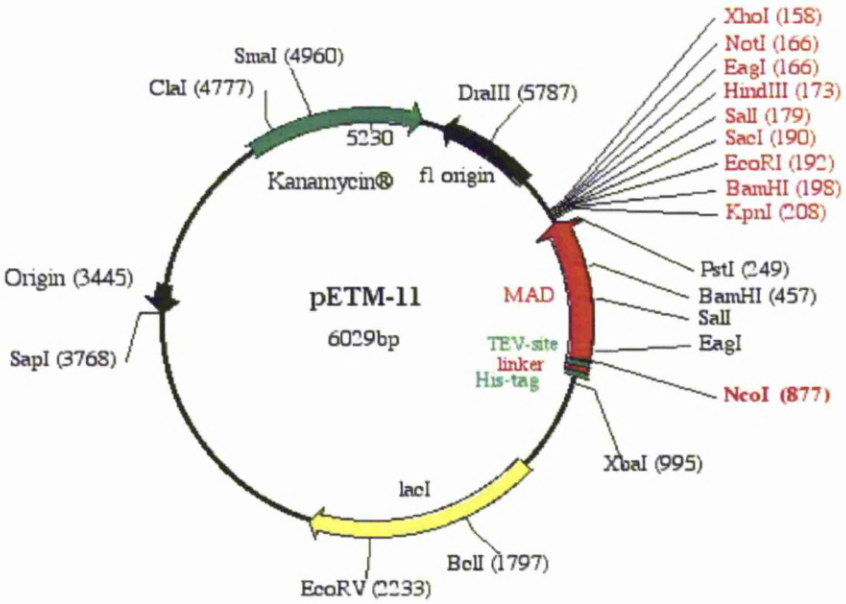


Figure 3.1. Vector map and multiple cloning site sequence of vector pET-M11.

3.3.1 Standard Cloning – pET-M11

For cloning into pET-M11 primers were designed complimentary to the AKAP sequences; the forward sequence included the restriction site for Nco1 and the reverse sequence included a stop codon followed by the Kpn1 restriction site sequence. The melting temperatures were similar at approximately 60°C.

PCR of primers for each reaction amplified with designed primers using the human AKAP79 gene supplied by Dr Caroline Dart, University of Liverpool, as a template and KOD hotstart polymerase (Novagen). PCR conditions were initial melting at 98°C for 2 minutes followed by 30 cycles of 95°C/15 seconds (DNA denaturation), 45°C/15 seconds (primer annealing) and 72°C/40 seconds (polymerase extension). The annealing temperatures were changed to match the melting temperature of each set of primers. PCR products were cleaned with QIAquick PCR purification kit (QIAGEN). PCR products and pETM-11 plasmid were digested by restriction enzymes, Nco1 and Kpn1 (New England Biolabs). Digest products were ligated with QuickLigase (Novagen) and transformed into XL-1 cells (Novagen). Inserts were verified by automated sequencing (Geneservice).

3.3.2 Ligation Independent cloning – pOPIN-S

Primer design differed with Ligation Independent cloning (Lic). The primers were still complimentary with the target sequence but including 15 bp extensions

homologous to the vector, for pOPIN-S 5'-GCGAACAGATCGGTGGT-3' and 3'-ATGGTCTAGAAAGCT-5'.

The PCR conditions used were the same as for pETM-11 above. The PCR product was incubated with DPN1 (New England Biolabs) to remove any remaining template and the PCR product cleaned with QIAquick PCR purification kit (QIAGEN). The pOPIN-S vector was cut with HindIII and NcoI (New England Biolabs) and cleaned before being mixed with PCR product in 1:1.5 vector to PCR product e.g. 200 ng/l cut vector + 300 ng/l PCR product made up to 10 µl with mili Q water. The PCR mix was then added to In-Fusion™ (Clonetech) dried reaction mix and incubated for 30 minutes at 42°C using a thermocycle. The reaction was stopped by the addition of 40 µl cold TE buffer and 4 µl of the ligated mixture was then transformed into 50 µl XL1 blue cells, heat shocked for 45 seconds at 42°C. 450 µl GS96 broth (for anaerobic bacteria) was added and incubated for 1 hour at 37°C. The cells were spun down (2 minutes 11000 g), 350 µL of broth removed and resuspended (by flicking) before the whole suspension is plated out and incubated overnight at 37°C. Control plates of 2 µl PCR product plus 8 µl mili Q water were also used. The In-Fusion™ Lic confers blue white selection that required agar plates containing 34 mg/l Kanamycin, 1 mM IPTG, 250 mg/ml X-gal (dissolved in 1mL DMF). The transformed colonies containing the insert DNA showed up white.

3.4 Protein purification

3.4.1 Calmodulin – unlabelled and labelled.

The calmodulin, CaM pET-15b vector was a gift from Dr. A.Kitmitto, Manchester University. Unlabelled CaM was prepared by transforming *E.Coli* BL21(DE3) from Novagen with the pET-15b CaM plasmid. Two microlitres of plasmid was added to 50 µL competent BL21 (DE3) cells, mixed by flicking and then incubated at on ice for 30 minutes. Cells were then heat shocked at 42°C for 45 seconds before being placed back on ice. Luria broth (LB) was then added up to 1000 µl and then incubated at 37°C for 60 minutes, shaking at 200 rpm. One hundred microlitres of the transformed cells were then aliquoted onto an antibiotic containing agar plate (100 mg/l ampicillin, Amp for CaM), spread evenly and incubated overnight at 37°C. The agar plate was removed for storage at 4°C for 1 to 2 weeks.

Cells were expressed at 37°C until 0.7 – 0.9 OD_{600} then induced with 1 mM IPTG for 3 hours at 37°C, harvested by centrifugation 11000 g for 15 minutes and resuspended in CaM lysis buffer (50 mM Tris, 2 mM DTT, pH 7.5). Cells were disrupted using the French press (Sim Aminco) at 1000 PSI and centrifuged (48000 g at 4°C for 30 minutes), the supernatant was recovered and $CaCl_2$ added to a final concentration of 2 mM. The supernatant was then heated to 65 °C for 3 minutes in a water bath with constant shaking before chilling on ice. The supernatant was then centrifuged once more (48000 g at 4°C for 20 minutes) to remove denatured

proteins. The resultant supernatant was collected and filtered through 0.22 μm syringe filter ready for chromatography.

The first purification step used hydrophobic interaction for separation. The supernatant was loaded onto a 20 mL HiPrep^{TM} phenyl sepharose column (GE Healthcare), pre-equilibrated with CaM buffer A (50 mM Tris, 200 mM NaCl, 2 mM $CaCl_2$, 0.5 mM DTT, pH 7.5). The column was washed with CaM buffer A, 3 column volume (CV) at 2 ml/min, followed by buffer B (50 mM Tris, 0.5 mM $CaCl_2$, 0.5 mM DTT, pH 7.5), 3 CV at 2 ml/min. The CaM was then eluted with 3 CV CaM buffer C (50 mM Tris, 1 mM EGTA, 0.5 mM DTT, pH 7.5). Eluted peak was concentrated to 5 ml using Amicon® Ultra (Millipore, 3K MWCO).

Secondly a gel filtration step was used to remove any remaining protein contaminants using 26/60 Superdex^{TM} 75 (GE Healthcare), pre-equilibrated with CaM gel filtration (GF) buffer (50 mM Tris, 200 mM NaCl, 10 mM $CaCl_2$, 0.5 mM DTT, pH 7.5). The sample was loaded before isocratic elution with 1 CV GF buffer at 1mm/min flow rate. The purity of the sample was assessed by SDS-Page gel and deemed to be >95% pure. CaM was concentrated using an Amicon® Ultra (Millipore, 3K MWCO) and then dispensed into microcentrifuge tubes and lyophilised before storage at -20°C.

To prepare uniformly ^{15}N labelled CaM, media containing ^{15}N ammonium chloride in a 2xM9 minimal media (MM) was used (See appendix 1 for details of media). The overnight step started with 1 colony grown in 1 ml LB for 4 hours before

removal of the LB and resuspension in 40 ml MM overnight at 37°C. The cells were centrifuged at 5000 g for 10 minutes to remove media before the whole cell pellet was added to 1 litre of MM. It was necessary to increase the cell density before adding the cell pellet to the MM or no growth would occur; however, this had to be balanced against exposure of the cells to LB, the greater the exposure the greater the decrease in isotopic labelling of the protein.

3.4.2 Kir2.1 C368 – labelled and unlabelled

The C-terminal cytoplasmic tail of Kir2.1 contains the PDZ binding motif ESEI. The C368 construct contains the last 60 residues at the C-terminus in a pGEX-6P-1 vector that contains a GST affinity tag that can be cleaved by the enzyme Precission protease (PP).

C368 was transformed into BL21 (DE3) RP^+ cells (Novagen) using the standard method above. For expression, cells were grown in LB plus 100 mg/l Amp to 0.7 – 0.9 OD_{600} before the cells were cold shocked by cooling flasks to approximately 4°C for 5 – 10 minutes before cells returned to 37°C to re-equilibrate. Cells were induced with 1 mM IPTG for 3 hours at 30°C, harvested by centrifugation for 15 minutes at 11000 g, and the cell pellets resuspended in PBS and frozen at -20°C overnight. The cells were thawed and disrupted by French Press at 1000 PSI, DNase1 (250 units) added before centrifugation (48000 g at 4°C for 30 minutes). The resultant supernatant was collected and filtered through 0.22 μm syringe filter ready for chromatography.

The affinity media used was Glutathione Sepharose^{TM} 4B (GE Healthcare). The resin has a binding capacity of approx 10 mg protein per ml resin. Resin is initially washed with low salt buffer to remove residual 20% ethanol storage buffer. The purification used a batch method. C368 supernatant was added to 2 ml Glutathione Sepharose 4B resin and incubated at 4°C for 60 minutes with shaking. The resin suspension was then decanted into a column, washed with 10 CV PBS to remove any residual unbound proteins and then equilibrated for the enzyme cleavage with 10 CV Precission Protease (PP) cleavage buffer. The resin was resuspended in 5 CV PP buffer and 1 mg PP was added and incubated overnight at 4°C. The resin was returned to the column and cleaved protein eluted, a further 5 CV PP buffer was used to elute any remaining protein. Eluted C368 was concentrated using an Amicon® Ultra (Millipore, 3K MWCO) and buffer exchanged using PD10 desalting column (GE Healthcare) to remove NaCl.

The sample was then further purified using a 1ml Mono Q 5/5 anion exchange column equilibrated with 10 CV anion exchange buffer (20mM Tris, 0.2mM DTT, 1mM EDTA, 1mM benzamidine HCl, pH 7.5) followed by 5 CV elution buffer (Anion exchange buffer + 1M KCl) and then a further 10 CV anion exchange buffer. C368 sample was loaded, washed and then eluted with a 0 – 500 mM KCl gradient over 20 CV at 1 ml/min flow rate. Peak samples were collected.

C368 is unstructured and thus more prone to proteolysis. To protect against degradation the purified protein was heat treated at 90°C for 10 minutes.

To prepare uniformly ^{15}N labelled C368 the same method of expression as with CaM above was used. The purification of,^{15}N C368 was performed identically to the unlabelled preparations.

3.5 Proteases

All the vectors used to express and purify the different proteins used affinity tags to aid with purification (except CaM). Each of these tags required removal. The vectors used contained specific restriction sites for highly specific enzymes. Three proteases were used, Precission protease for pGEX plasmids, Tev protease for pETM-11 and pLEICS plasmids and SUMO protease for pOPIN-S plasmids. The proteases also contain the same affinity tag that they will cleave but without the cleavage sequence. This helps in purifying the protease and once the protease has removed the tag the resultant mixture can be passed back down the affinity resin and only the cleaved protein will be eluted.

3.5.1 Precission protease purification

Precission protease was transformed into BL21 (DE3) RP^+ cells (Novagen) using the standard method above. For expression, cells were grown in LB plus 100 mg/l Amp to 0.7 – 0.9 OD_{600} before the cells were cold shocked by cooling flasks to approximately 4°C for 5 – 10 minutes before cells returned to 37°C to re-

equilibrate. Cells were induced with 1 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation for 15 minutes at 11000 g, the cell pellets resuspended in PBS and frozen at -20°C overnight. The cells were thawed and disrupted by French Press at 1000 PSI, DNase1 (250 units) was added and centrifuged (48000 g at 4°C for 30 minutes). The resultant supernatant was collected and filtered through 0.22 μm syringe filter ready for chromatography.

The PP plasmid contains a GST tag; hence, Glutathione Sepharose™ 4B (GE Healthcare) was used. The batch method was used for the purification. PP supernatant was added to 5 ml Glutathione Sepharose 4B resin and incubated at 4°C for 60 minutes with shaking. The resin suspension was then decanted into a column, washed with 10 CV PBS to remove any residual unbound proteins and then equilibrated with 10 CV PP cleavage buffer. The PP was eluted with 3 CV PP buffer + 30 mM reduced glutathione (Sigma). Eluted PP was concentrated using an Amicon® Ultra (Millipore, 10K MWCO).

This was followed by a gel filtration step using a 26/60 Superdex™ 75 (GE Healthcare), pre-equilibrated with PP buffer. The sample was loaded before isocratic elution with 1 CV GF buffer at 1 mm/min flow rate. The PP was concentrated with an Amicon® Ultra (Millipore, 10K MWCO) to 1 mg/ml and 1 ml aliquots flash frozen in liquid nitrogen before storage at -80°C

3.5.2 Tev Protease purification

Tev protease was transformed into BL21 (DE3) cells (Novagen) using the standard method above. For expression cells were grown in DYT plus 50 μg/l Kan to 0.7 – 0.9 OD_{600}. Cells were induced with 1 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation for 15 minutes at 11000 g. The cell pellets were resuspended in His buffer (0.5 M NaCl, 50 mM Tris-HCl, 2 mM DTT, pH 8) and frozen at -20°C overnight. The cells were thawed and disrupted by French Press at 1000 PSI, DNase1 (250 units) was added and centrifuged (48000 g at 4°C for 30 minutes). The resultant supernatant was collected and filtered through 0.22 μm syringe filter ready for chromatography.

Tev protease contains a hexhistadine tag for purification. The supernatant was bound to a 5ml HisTrap™ Fast Flow (GE Healthcare), washed with increasing concentrations of imidazole in His buffer (10, 20, 50 mM), at 2 ml/min flow rate, 6 CV at each increment, then eluted with 250 mM imidazole in His buffer. Eluted peak was buffer exchanged using HiPrep™ 26/10 desalting column (GE Healthcare) into tev cleavage buffer (150 mM NaCl, 50 mM Tris HCl, 2 mM DTT pH 7.5) concentrated with an Amicon® Ultra (Millipore, 10K MWCO) to 1 mg/ml and 1 ml aliquots flash frozen in liquid nitrogen before storage at -80°C.

3.5.3 Sumo Protease Purification

SUMO protease was expressed and purified using the same methods as Tev protease, with the only difference being the addition of a second purification step after the nickel affinity step (i.e. post HisTrap^{TM}).

The SUMO protease was buffer exchange using a HiPrep^{TM} 26/10 (GE Healthcare) desalting column to remove the imidazole, before a further purification using a 5 ml HiTrap^{TM} Q FF anion exchange column equilibrated with 10 CV anion exchange buffer (20 mM Tris, 2 mM DTT, 1 mM EDTA, 1 mM benzamidine HCl, pH 7.5) followed by 5 CV elution buffer (Anion exchange buffer + 1 M KCl) and then a further 10 CV anion exchange buffer. Sumo Protease sample was loaded, washed and then eluted with a 0 – 500 mM KCl gradient over 20 CV at 2 ml/min flow rate. Peak samples were collected concentrated with an Amicon® Ultra (Millipore, 10K MWCO) to 1 mg/ml and 1 ml aliquots flash frozen in liquid nitrogen before storage at -80°C.

3.6 Experimental Methods

3.6.1 SDS Page Gel Electrophoresis

A 0.75 mm glass mould was used with a 15 % resolving gel (50% of 30% acrylamide, 1.5 M Tris pH 8.8, dH_2O, 0.1% SDS, 100 mg/ml ammonium persulphate (APS) and 0.1% TEMED) and 4% stacking gel (13% of 30%

acrylamide, 0.5 M Tris pH 6.6, dH_2O, 0.1% SDS, 100 mg/ml APS and 0.1% TEMED) before a comb was inserted. Five micro litres of sample was then diluted with 45 µl dH_2O then 50 µl 2x sample buffer was added (0.5 M Tris, 50% glycerol, 1.6% SDS, 4% β-mercaptoethanol and 0.1% bromophenol blue, pH 6.8). The samples were heated for 5 minutes at 100°C. Ten microlitres of sample was then loaded into each lane with 5 µl of low range molecular markers (Sigma) also loaded per gel.

Gels were ran using Biorad Protean II kit at 200 V for approximately 45 minutes and then stained for 1 hour in Coomassie G250 (0.1% Coomassie blue G250, 45% methanol, 45% dH_2O, 10% acetic acid). The gel was then destained overnight in dH_2O.

3.6.2 Dansylation of Calmodulin

The dansylation of calmodulin followed a previously published method (Vorherr et al 1990). Purified CaM was diluted to 50 µM in 20 mM ammonium bicarbonate, pH 7.5 and Ca^{2+} added to a concentration of 1 mM. The CaM was then mixed 1:1 with 50 µM dansyl chloride in acetone. The mixture was incubated at room temperature for 2 hours and vortexed every 20 minutes. After incubation the mixture was dialysed (3 KD MWCO) in 4 litres 20 mM ammonium bicarbonate, pH 7.5 overnight 4°C. The dansyl CaM (d-CaM) was recovered, centrifuged to remove any particulates then buffer exchanged into fluorescence buffer (20 mM Tris, 150 mM NaCl, 2 mM $CaCl_2$ pH 7.5) and concentrated to 50 µM d-CaM, using Amicon® Ultra (Millipore, 3K MWCO).

3.6.3 Fluorimetric Experiments

Fluorescence measurements were performed on a Perkin Elmer fluorometer. Excitation was at 340 nm and emission measured between 400 – 600 nm, through a slit width of 5 nm, curve smoothing was also applied. A 200 µL fluorometer cuvette with a path length 1 cm was used for all measurements. SAP97 Δ70, Δ461 and Δ546 were buffer exchanged into fluorescence buffer and concentrated to 200 µM. The SAP97 constructs were then titrated into 200 µl 5 µM d-CaM in 0.5 µl aliquots until saturation occurred. Saturation was observed when three readings in succession were taken without increase in measured intensity. Each SAP97 construct was repeated three times. Buffer blanks were then taken where fluorescence buffer was titrated into 5 µM d-CaM, three replicates were taken. Analysis measured SAP97 construct concentration minus the buffer blank against normalised fluorescence intensity. A plot was derived using Sigma plot 11.0, one site saturation curve, which generated a K_d and standard error for each titration.

3.6.4 NMR Experiments.

$^{13}C^{15}N$ Calmodulin (0.2 M) was titrated with unlabelled synthetic $AKAP79_{31-52}$ from a ratio 1:0.5 to 1:10 containing 50 mM Tris, 5 mM $CaCl_2$, pH 6.5, 295K. The pH was checked for each titrant and experiments ran at 600 MHz using 5 mm O.D Wilmard tubes (Sigma). 1D and 2D ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence) data acquired with instruction from Dr Marie Phelan using the programme Topspin (Bruker).

All other NMR experiments were performed at either 800 or 600 MHz, using 5mm O.D. NMR Wilmard or Shigemi tubes (Shigemi Inc) and were acquired by Professor Lu-Yun Lian. For AKAP79 M assignment 2D ^{1}H-^{15}N HSQC and 3D CBCANNH and CBCA(CO)NNH experiments were used and data analysed using CCPNMR Analysis, this will be explained in more detail in following chapters.

3.6.5 Size Exlusion Chromatography – Multi-angle Laser Light Scattering, SEC-MALLS

All experiments were performed by the Biomolecular Interactions Analysis Facility (BMIF) at the University of Manchester using a Dawn DSP (Wyatt Technology). The scattering data was then processed using Dynamics^{TM}. The samples were concentrated to 1 mg/ml and centrifuged to remove particulates before 200 µl of sample applied to the machine.

3.6.6 Circular Dichroism, CD

Circular Dichroism, CD is a technique used to determine the secondary structure and the folding properties of proteins. The technique utilizes the ability of different structures to absorb right and left handed circular polarised light to different extents (Greenfield, 2006).

All experiments were performed by the Biomolecular Interactions Analysis Facility (BMIF) at the University of Manchester using a Jasco J-810 (JASCO UK) with 150 W Xenon lamp. The CD data was then processed using Origin^{TM} software. The

lyophilised samples were resuspended in 50 mM Tris pH 7.0 to 1 mg/ml and centrifuged to remove particulates before inserting into a 100 µl CD cuvette. A simple titration series with increasing concentrations of trifluoroethanol acid (TFE) was also carried out. The percentage α-helical formation was calculated from the equations

$$\%\text{helix} = 100 \times (\theta_{obs} - \theta_C)/ (\theta_H - \theta_C)$$

$$\theta_C = 640 - 45T$$

$$\theta_H = -40000 \times (1 - \chi / N) + 100T$$

where θ_{obs} is the observed molar elipticity at 222 nm in deg.cm^2/dmol, T = temperature in °C, χ is a constant set at 2.5 and N is the number of amino acids in the peptide (Scholtz, Hong et al. 1991)

3.6.7 Isothermal Titration Calorimetry, ITC

Protein – protein interactions can be calculated by Isothermal Titration Calorimetry (ITC). Minute changes in enthalpy, either endothermic or exothermic can be measured by ITC. The instrument has two cells, one to hold sample, the other a reference cell. A syringe containing the binding sample is inserted into the sample cell. With each aliquot the temperature difference between the sample cell and the reference cell is measured. This is achieved by measuring the amount of heating or cooling that is required by the sample cell jacket to match the reference cell. Once both cells are back to the same temperature another aliquot is added. The data generated is processed using the built in Origin™ software.

ITC requires only small volumes, approximately 70 µl ligand in the syringe and 300 µl protein in the sample cell per run. Due to the sensitive nature of the technique both protein and ligand were dialysed overnight in the same buffer using Slide-A-Lyzer Dialysis Cassettes (Pierce). Both ligand and protein were then concentrated by Amicon® Ultra (Millipore). The protein was carefully inserted into the sample cell avoiding bubbles and the ligand taken up into the syringe. The syringe was then placed into the sample cell and spinning started. When the temperatures in the sample and reference cells were equal the titration was started. Control titrations of buffer into buffer, ligand into buffer and buffer into protein were also tested.

3.6.8 Small Angle X-ray Scattering, SAXS

The SAXS data were collected from the Soleil synchrotron, Gif-sur-Yvette, France using the SWING beam line with the help of beam line scientist Dr. Javier Perez. The scattering range, S was set to 0.07 – 0.5 $Å^{-1}$, where $S = 4\pi\sin\theta / \lambda$, 2θ is the scattering angle and $\lambda = 1.03$ Å (X-ray length). The beam size was 0.4 x 0.1 mm and the flux was about 10^{12} photons $second^{-1}$ (David and Perez, 2009). Samples were prepared at concentrations between 1 – 5 mg/ml with at least 150 mM NaCl (for HPLC only) and 5 mM DTT (to stop radiation damage), centrifuged to remove particulates then approximately 100 µl added to an Agilent HPLC vial with insert. The system set up at Soleil, Figure 3.2, utilises two methods. In the standard batch method a continuous stream of dH_2O at 40 µl/minute was passed through a capillary flow cell and through the X-ray beam. A 40 µl sample was then injected into the dH_2O stream flanked by two bubbles of air that the syringe inserted to

prevent the sample diluting and to enable visual recognition of when the sample was passing through the beam using the video camera that is always focussed on the flow cell. Approximately 250 frames of data were collected for each sample at three concentrations and this was repeated with the sample buffer for a control. The data was initially processed by Dr. Javier Perez using Primus, the frames were averaged for each concentration and the background control subtracted.

In the second method an HPLC system with a 700 kDa size exclusion column (Shodex, KW403 -4F, 4.6 mm inner diameter (ID) x 300 mm long) was used. 80 µl of each sample was loaded at 150 µl/minute flow rate and passed through an UV detector, and then through the X-ray beam. 255 frames over 13 minutes were collected with the flat baseline used as the background control. The data were then collated by the in house program Foxtrot. The frames responsible for each peak on the chromatogram were then collated and averaged and the background subtracted. These data were then processed as before by Primus.

The averaged experimental data was then processed by several other computer programs with differing algorithms including GNOM, GASBOR, BUNCH and EOM (see chapter 2 for more detailed information) to enable computation of R_g (radius of gyration), D_{max}, molecular envelopes and representative models of the proteins.

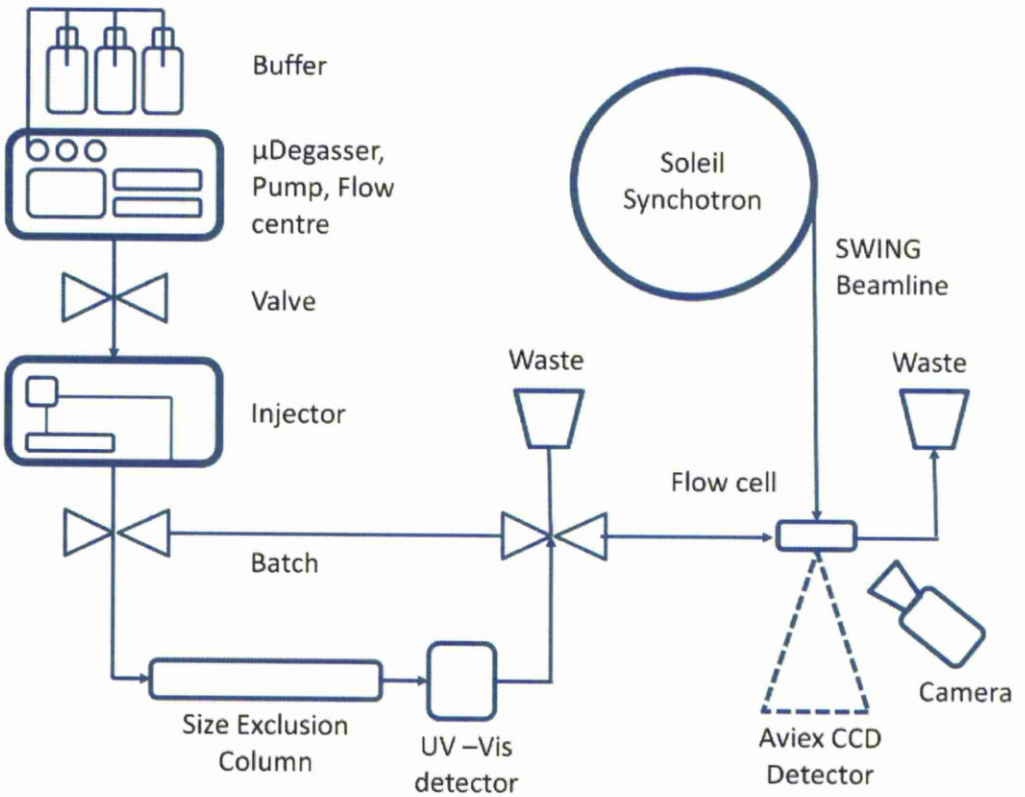


Figure 3.2. Flow diagram of Soleil SWING beam line. Inside the SWING hutch showing the possible path of both batch and HPLC methods. The sample vials are inserted into the injector port.

CHAPTER 4

CHARACTERISATION OF A KINASE ANCHORING PROTEIN, AKAP79

4.0 Introduction

There are three main aspects to each AKAP: a PKA anchoring domain, a means to compartmentalise the protein and the ability to bind or interact with other signalling molecules (Welling, 2008).

AKAP79 is 427 amino acids long in which several binding domains have been identified, Figure 4.0.1. At the N-terminal domain, three binding sites to the plasma membrane were identified (Dell'Acqua et al., 1998) and of these one was identified to be a calmodulin and PKC binding site (Faux and Scott, 1997). There is also evidence that the central domain contains a possible binding site to MAGUK proteins including SAP97 (Colledge et al., 2000). At the C-terminal there is a calcineurin binding domain and a PKA anchoring domain that has been the focus of much study (Klauck et al., 1996).

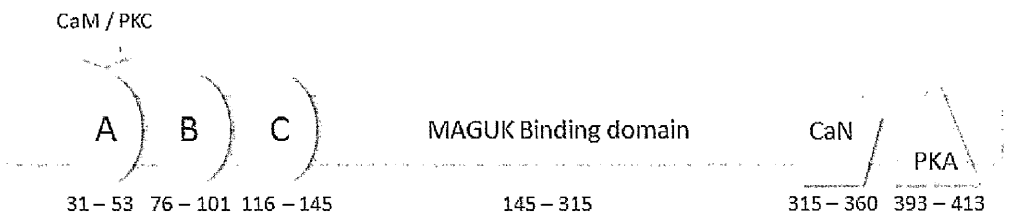


Figure 4.0.1. Schematic of AKAP79 showing the locations of binding sites. Binding sites A, B and C represent the three hydrophilic regions proposed to be membrane binding sites. Site A also binds calmodulin (CaM) and PKC. The exact MAGUK binding domain is unknown but has been shown to lie within the region shown (Colledge et al., 2000). The PKA and calcineurin sites are towards the N-terminus of the AKAP79 (Dell'Acqua et al., 1998).

Most of the work involving AKAP79 was carried out in Hek293 mammalian cell lines or using brain extracts. Thus little structural characterisation has been carried out.

With this in mind, the aims here are to examine and characterise the structural aspects of AKAP79 and individual domains of AKAP79 that were previously described (Dell'Acqua et al., 1998), Figure 4.0.2.

The four new AKAP79 constructs to be characterised were cloned into pETM-11, a vector created at EMBL derived from pET-24d with the inclusion of a TEV protease cleavage site between the hexahistag and the cloning site (Pinotsis et al., 2006). A summary of the constructs made are given in Table 4.0.

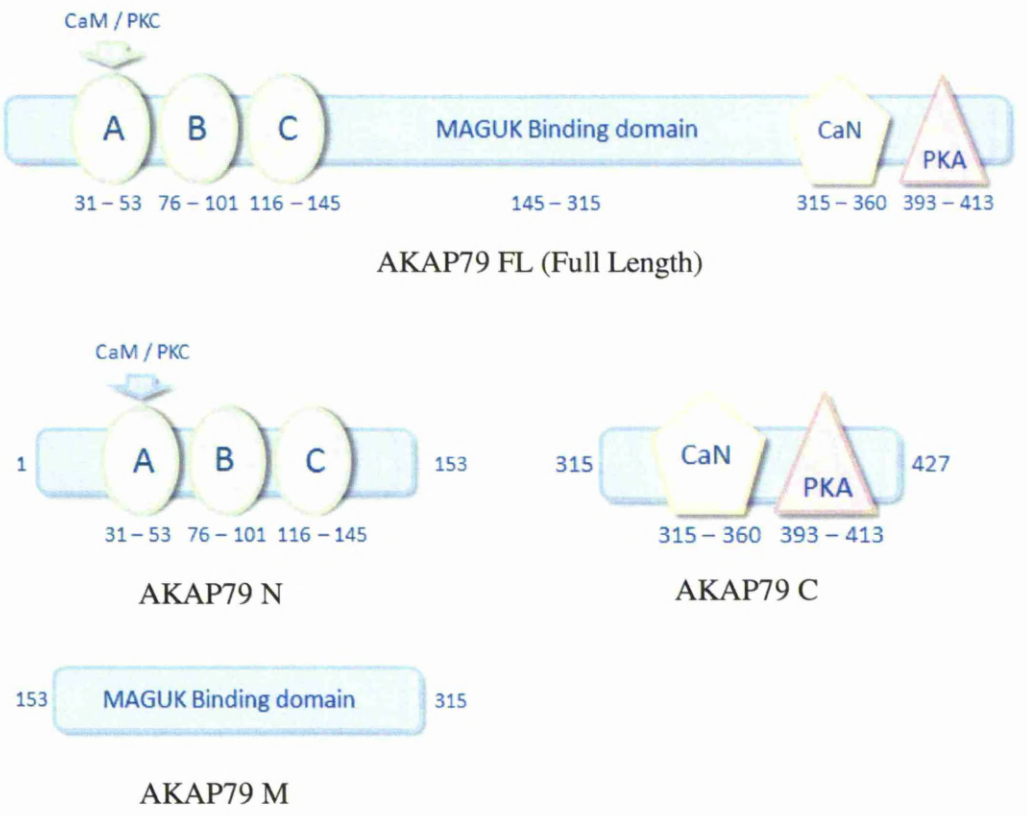


Figure 4.0.2. Schematic of AKAP79 domains. Each of the separate domains were created to characterise AKAP79 and its binding domains. Each domain was cloned into pETM-11 vector

Table 4.0 Summary of AKAP79 constructs. Including purification tags, specific cleavage site and vector to be cloned into.

Protein	Construct	Vector
AKAP79 Full Length Mr = 47 kDa	HexaHis-TEV-1 – 427	pETM-11
AKAP79 N Mr = 21.7 kDa	HexaHis-TEV-1 – 153	pETM-11
AKAP79 M Mr = 21.5 kDa	HexaHis-TEV-153 – 315	pETM-11
AKAP79 C Mr = 15.9 kDa	HexaHis-TEV-315 – 427	pETM-11
AKAP79 M Sumo Mr = 30.7 kDa	HexaHis-SUMO cleavage-153 – 315	pOPIN-S

4.1 Results

During the course of the work a new cloning strategy was employed to introduce a fusion protein to protect the constructs from degradation and increase recombinant protein expression. Ligation independent cloning (LIC) is a high efficiency cloning strategy that was developed for high throughput screening of proteins. It forgoes the need for restriction enzymes and DNA Ligase, hence streamlining the procedure (Berrow et al., 2007). The LIC used was In-Fusion^{TM} developed by Clonetech. Primers are designed with a 15 bp extension to both forward and reverse primers, complementary to the vector to be used. It uses a custom enzyme with exonuclease and proofreading activities that can recognise and fuse the ends of any two complementary linear DNA fragments, thus enabling precise insertion into vectors (Marsischky and LaBaer, 2004). This creates a three step process; amplification of the primers, followed by fusion to the vector and finally transformation (Figure

4.1). The vectors used were pOPIN vectors designed by the Oxford Protein Production Facility (OPPF); these confer blue white screening of clones. The specific vector used was pOPIN-S which also contained a cleavable SUMO fusion protein containing a hexahistidine tag.

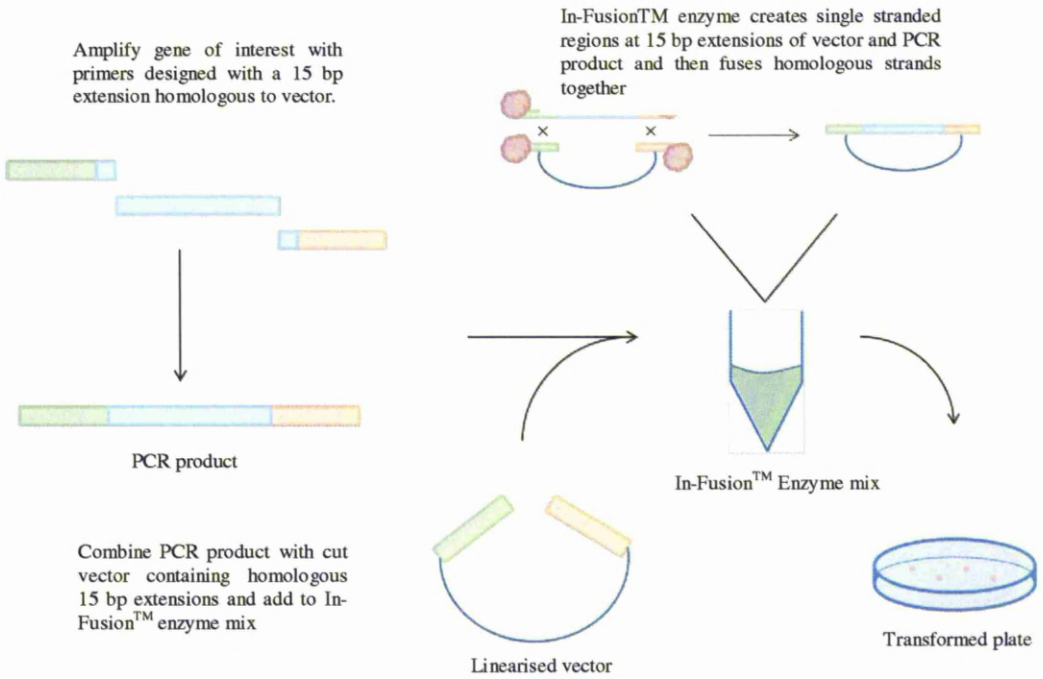


Figure 4.1. In-Fusion™ Cloning. The pictorial scheme outlines the steps used in In-Fusion™ Cloning. The gene is amplified with primers that contain a 15 bp extension by PCR. The product is then added with cut vector to the In-Fusion™ Enzyme mix and heated to 42 °C for 30 minutes for the fusion to occur before transformation onto an agar plate with IPTG and β-galactosidase for selection by Blue White screening. The figure was adapted from the In-Fusion™ Cloning manual produced by Clonetech.

4.1.1 AKAP79 pETM-11 Cloning, Expression and Purification

Primer for AKAP79 full length (FL) and each of the three domains (N, M and C) were designed, including restriction sites for KpnI and NcoI (NEB). The initial PCR gave good amplification of each construct (Figure 4.1.1). Insertion into the vector and ligation using Quick T4 DNA ligase (NEB) worked efficiently. Four colonies of each construct were then selected for screening by PCR using the initial primers (Invitrogen). All fours constructs AKAP FL, AKAP N, M and C domains gave strong bands inferring successful ligation Figure 4.1.2. The positive colonies were then grown up in 10 ml LB overnight and the plasmid extracted by mini-prep (Qiagen). Samples of each plasmid were then sent for Sanger sequencing (Geneservice) using stock T4 forward and reverse primers. The full sequencing confirmed the designed clones (see appendix 3).

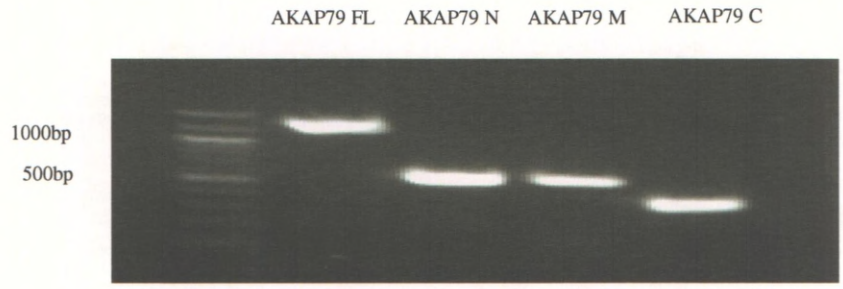


Figure 4.1.1. 1% Agarose gel depicting initial PCR of AKAP79 constructs. Each band shows that each construct is the correct size, AKAP79 FL is approximately 1300 bp, AKAP79 N and M 500 bp and AKAP79 C is 350 bp.

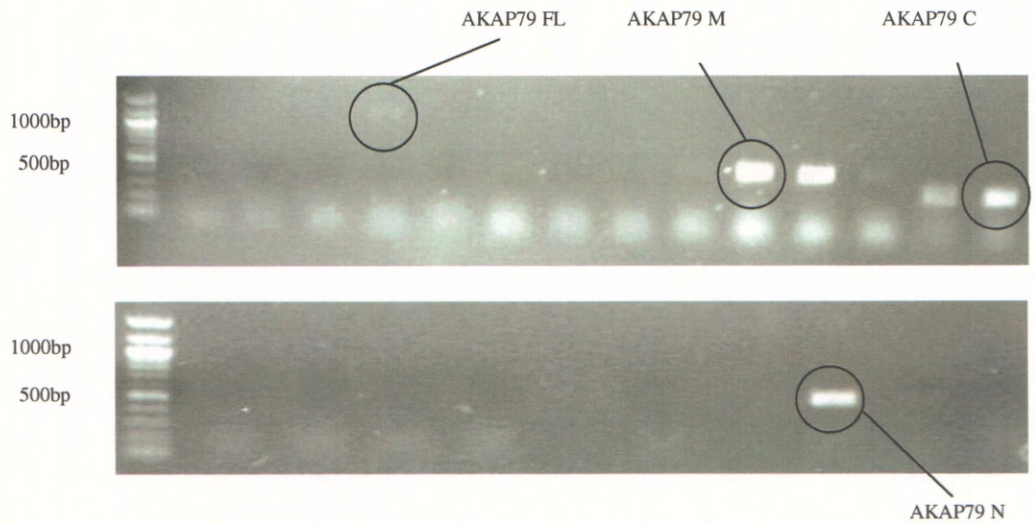


Figure 4.1.2. 1% Agarose gel depicting PCR screen of AKAP79 constructs post ligation and transformation. Transformed colonies were picked from agar plate, diluted in mili Q and used as a template for the PCR. The initial PCR primers for each construct were used for each reaction. AKAP79 FL, M, N and C gave positive results.

4.2 AKAP79 Full-length

4.2.1 Expression and Purification

AKAP FL pETM-11 vector was transformed in competent B834 cells containing the Takara pG-KJE8 plasmid. The pG-KJE8 plasmid contains the code for five bacterial chaperones, dnaK, dnaJ groEL, groES and grpE. The plasmid confers chloramphenicol resistance for plasmid selection and has a tetracycline (Tet) and L-arabinose promoter for induction of chaperone expression. The chaperones reduce inclusion bodies and increase soluble expression (Nishihara et al., 1998).

Unlabelled AKAP97 FL were expressed in 1 litre LB plus Kan/CM, 10 ng/mL Tet and 3 g/L L-arabinose to 0.7 – 0.9 OD_{600} at 37°C before the cells were cold shocked and induced with 200 μM IPTG overnight at 18°C. Cells were pelleted at 11000 g for 15 minutes and resuspended in His buffer (0.5 M NaCl, 50 mM Tris HCl, 2 mM β-mercaptoethanol (β-ME pH 8) plus one EDTA free protease inhibitor tablet (Roche) and 250 ng DNase I. Stored at -20°C overnight.

The cells were thawed and lysed by French press at 1000 PSI, then 10 mM $MgCl_2$ and 5 mM ATP (bacterial) was added before being incubated at 4°C for 30 min with mixing to remove any chaperones still bound to the AKAP79 FL. The lysed cells were centrifuged (48000 g at 4°C for 45 minutes). The resultant supernatant was collected and filtered through 0.22 μm syringe filter ready for chromatography.

AKAP79 FL was purified in a two step process. Firstly, it was bound to a 5ml HisTrap™ High Performance (GE Healthcare), washed with increasing concentrations of imidazole (10, 20, 50 mM), at 2 ml/min flow rate, 6 CV at each increment and then eluted with 250 mM imidazole. Eluted peak was buffer exchanged using PD10 desalting column to remove imidazole and concentrated to approximately 5 ml, ready for next purification step.

In the second step, the AKAP79 FL was further purified using a 1 ml Mono Q 5/5 anion exchange column. The sample was loaded, washed and then eluted with a 0 – 500 mM KCl gradient over 20 ml at 1 ml/min flow rate.

4.2.2 Results

4.2.2.1 Expression and Purification

AKAP FL was initially expressed in B834(DE3) cells but showed no expression at 37°C, 30°C and 18°C overnight induction with either 1mM or 0.2mM IPTG (Melford). BL21(DE3) RP$^+$ and Rosetta™(DE3) cells were then tested with Rosetta™(DE3) only showing very weak expression. It was hypothesised that because AKAP79 had very little tertiary structure the bacterial host cells were degrading the protein by proteolysis. To solve this, expression using B834(DE3) cells containing the Takara plasmids were tried. The best results were obtained using cells containing the Takara pG-KJE8 plasmid, which expresses the chaperones dnaK, dnaJ groEL, groES and grpE, in LB media. A band of an approximate size of 66 kDa was seen by SDS PAGE, Figure 4.2.3.

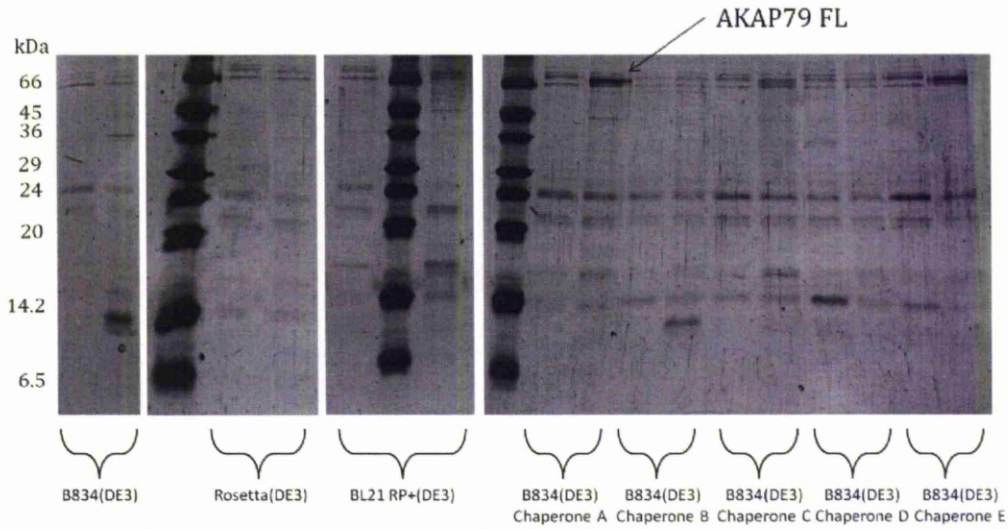


Figure 4.2.3. Expression tests for AKAP79 FL. Each sample was induced with 0.2mM IPTG at 18°C for approximately 18 hours. The samples were lysed using Bugbuster™ (Pierce) centrifuged at 48000 g and supernatants applied to prewashed Nickel affinity resin (GE). Resin was washed with 10 mM imidazole before the resin was loaded onto the SDS PAGE gel. For each different strain the first lane is the pre-induced followed by the imidazole elution that shows soluble protein. Chaperone A contained dnaK, dnaJ, grpE, groES and groEL. Chaperone B groES and groEL, Chaperone C dnaK, dnaJ and grpE. Chaperone D groES, groEL and tig. Chaperone E tig only.

Cell lysis was performed using a French press (Sim Aminco) and the resultant lysed cells centrifuged for 30 minutes at 48000 g to remove cell debris. The resultant fluid phase appeared in two layers; the expected clear yellow tinged solution (supernatant) and a second more dense brown layer (loose pellet). Increasing the centrifugation time to 45 minutes helped to reduced the presence of the loose pellet and successfully pelleted the solid phase. Analysis of the loose pellet showed that it did contain 5-10% AKAP79 FL together with other *E.Coli* proteins. To reduce the loose pellet further the reducing agent DTT (Melford) was changed to β-mercaptoethanol (Sigma), the rationale behind this was that the chaperones dnaK, dnaJ contained zinc fingers that precipitated when reduced by DTT

The cell lysate was purified on a Ni^{2+} Affinity column, (5 ml HisTrap HP, GE) Figure 4.2.4. After elution with 250mM imidazole, the SDS PAGE gel showed the presence of AKAP79 together with other contaminating proteins after affinity chromatography. To decrease the level of the contaminating proteins, 5 mM Mg^{2+} and 5 mM ATP were added to the cell lysate as together they are known to induce a conformation change in DnaK that lowers its substrate binding affinity (Guo et al., 2007).

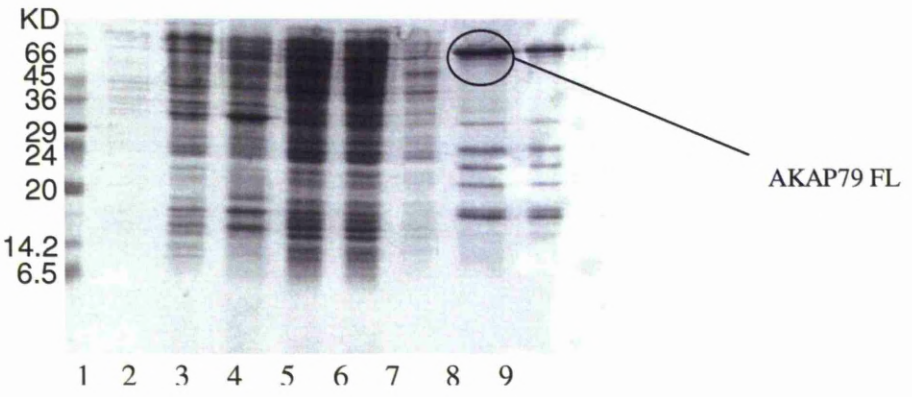


Figure 4.2.4. AKAP79 purified on 5ml Histrap HP column (GE). SDS PAGE Gel shows the peak fraction after addition of 250 mM imidazole. The prominent band at approximately 66 kDa is AKAP79, the other bands are impurities, most likely bacterial proteins or Takara chaperones. Lanes 2 through to 9 are pre(2) and post(3) induction, lysate(4), and pellet(5) following French press, Ni^{2+} affinity flowthough (6) and buffer wash(7) and 250 mM imidazole elution (8) and elution post buffer exchange (9).

The AKAP FL from the Ni^{2+} affinity column was pooled, buffer exchanged by PD10 (GE) and further purified using a Mono Q ion exchange column (1 ml MonoQ 5/50 GL, GE) Figure 4.2.5. The SDS PAGE showed that the main peak

was found in fractions 5, 6 and 7; fractions were pooled and concentrated to to yield approximately 4 mg AKAP79 per litre LB.

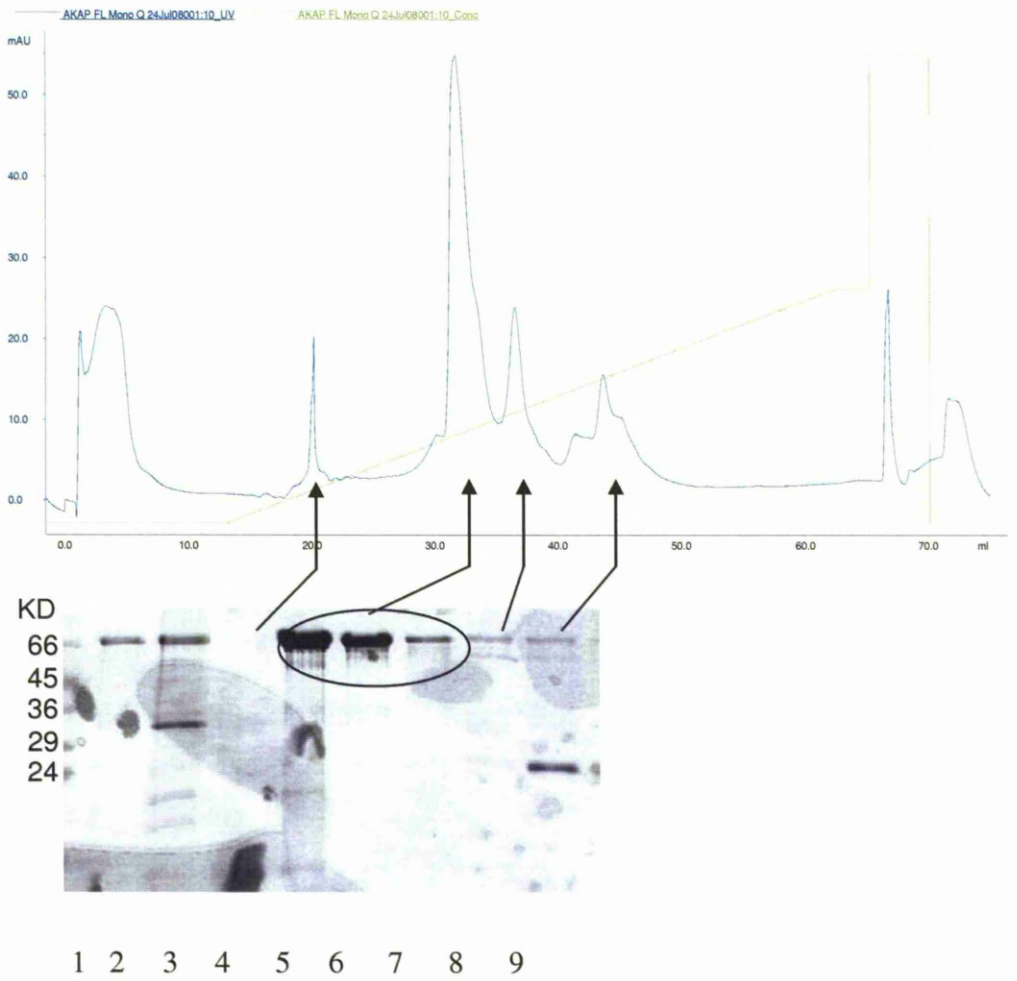
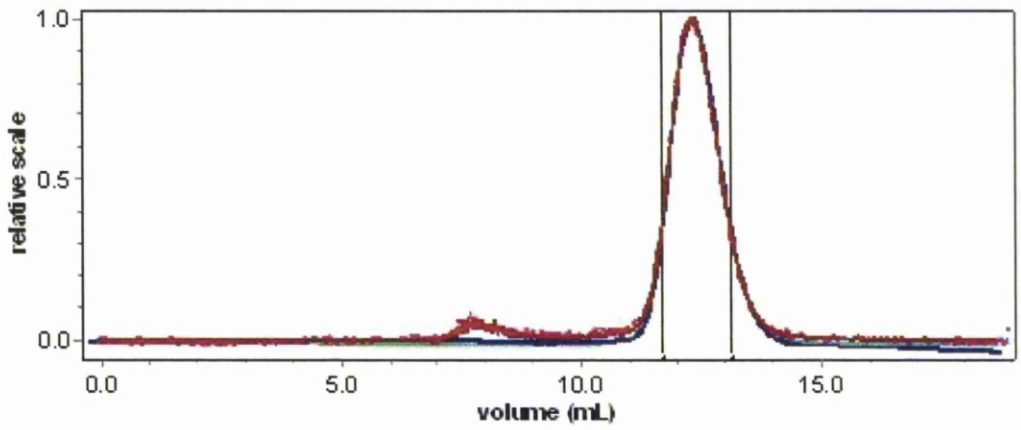


Figure 4.2.5. AKAP79 purified on 1 ml MonoQ 5/50 GL (GE). Mono Q was run at a flow rate 1 mL/min with a gradient 0 – 500 mM KCl over 40 column volumes. Lane 1 shows Low range markers, Lane 2 sample post HisTrap, Lane 3 concentrated sample post PD10 that also increases the concentration of the contaminants and Lanes 4-9 Mono Q elution. SDS PAGE Gel insert shows pure protein in the main peak that corresponds to AKAP79 at approximately 66 kDa.

The gel bands of the 66 kDa band were cut out and sent for identification by mass spectrometry and confirmed AKAP79 FL.

4.2.2.2 SEC-MALLS and NMR

A sample of the purified AKAP79 was subjected to size exclusion chromatography – multi angle laser light scattering, (SEC-MALLS, Dawn DSP, Wyatt Technology) Figure 4.2.6. The scattering data was processed using Dynamics^{TM} and provided a molecular mass of 43 kDa and a hydrodynamic radius of 6.6 nm. The molecular mass obtained was smaller than the expected 45 kDa but this type of result is normal for an unfolded protein. Another observation was the eluate was not initially homogenous perhaps due to a certain degree of aggregation so the main peak was collected and ran again through the SEC-MALLS. The Hydrodynamic radius was also greater than expected for a 45 kDa protein. This was also most likely due AKAP79 full length being unfolded.



Molar mass moments (g/mol)		Hydrodynamic radius moments (nm)	
Mn	4.290e+4	Rh(n)	6.6
Mp	4.304e+4	Rh(w)	6.6
Mw	4.297e+4	Rh(z)	6.6
Mz	4.304e+4	Rh(avg)	2.1
M(avg)	4.287e+4		
rms radius moments (nm)		Translational diffusion moments (cm^2/sec)	
Rn	n/a	Dt(n)	9.49e-7
Rw	n/a	Dt(w)	9.46e-7
Rz	n/a	Dt(z)	9.43e-7
R(avg)	n/a	Dt(avg)	3.55e-8

Figure 4.2.6. AKAP79 full length subjected to SEC-MALLS. The data collected gave an average molecular weight of 43 kDa and a hydrodynamic radius of 6.6. Green line shows absorbance, blue line the refractive index and red line the light scattering.

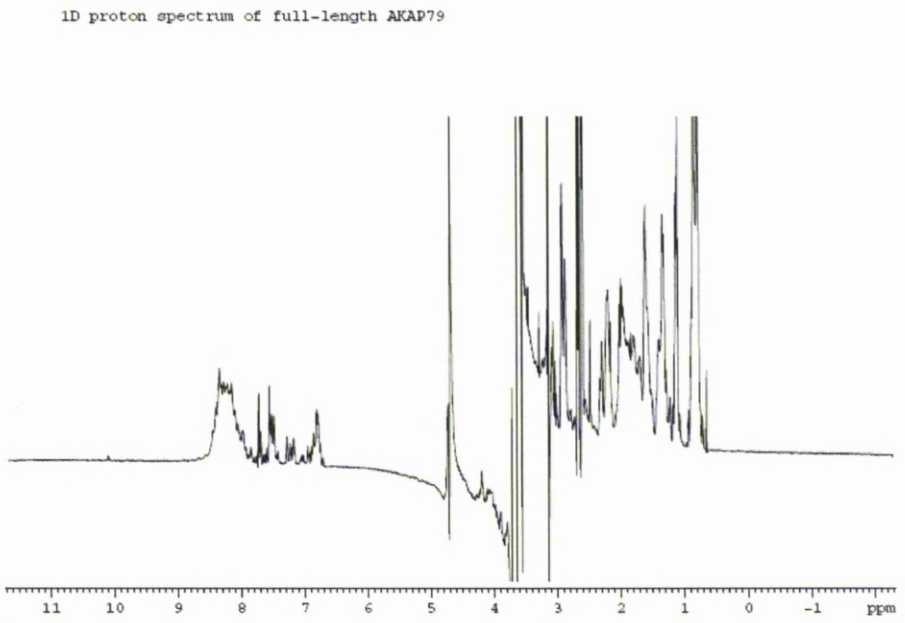


Figure 4.2.7. 600 MHz 1D ^{1}H-NMR protein spectrum of unlabelled AKAP79 FL in 50 mM phosphate, 50 mM NaCl, 0.01% NaN_3 pH 7.5 at 298 K.

The ^{1}H NMR data shows that full length AKAP79 (Figure 4.2.7) is an unfolded protein as there is little dispersion of the amino acids resonances. This may explain why attempts to crystallise the protein have so far failed to yield successful results. The lack of dispersion in the amide region is revealed more clearly in the ^{15}N-^{1}H HSQC spectrum shown in Figure 4.2.8.

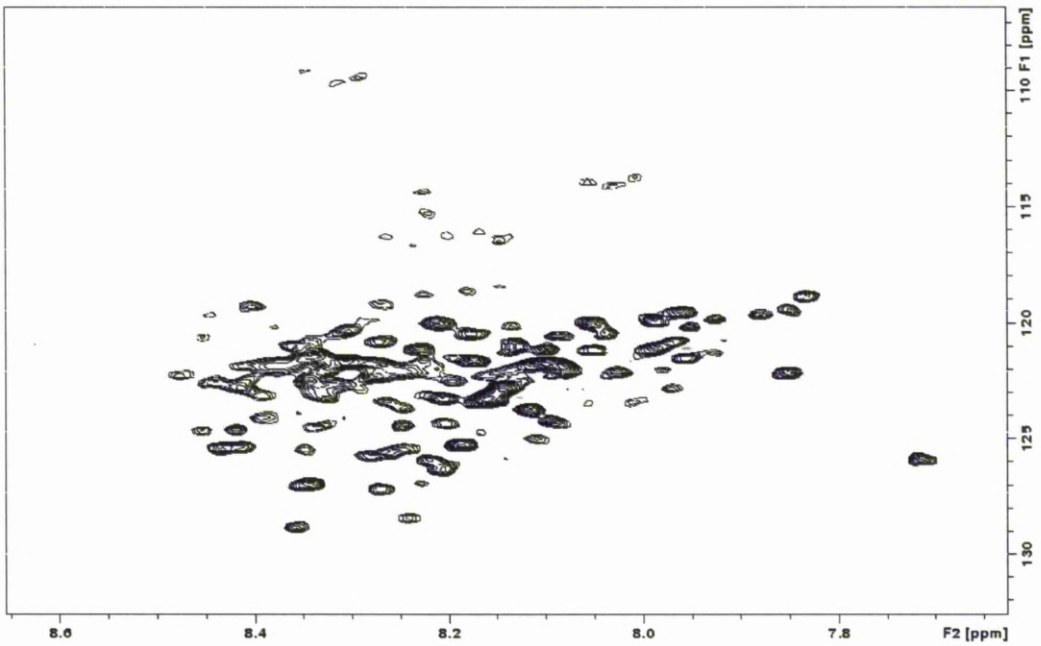


Figure 4.2.8. 600 MHz 2D ^{15}N-^{1}H HSQC spectrum of ^{15}N AKAP79 FL in 50 mM phosphate, 50 mM NaCl, 0.01% NaN_3 pH 7.5 at 298 K. Peaks are not dispersed evenly showing that the protein is unfolded.

4.3 AKAP79 N and C Domains

Expression of these domains from the pETM11 constructs were screened against all the same bacterial host strains - BL21 RP^{+}(DE3), Rosetta^{TM}(DE3) and B834(DE3) including the TAKARA chaperone plasmids- as previously described with AKAP79 FL. Unfortunately, no bacterial strain gave any detectable expression. This was possibly because of a higher level of proteolysis of the proteins during expression due to their shorter lengths. Due to time constraints, no further work was undertaken for these domains

4.4 AKAP79 M

4.4.1 Cloning

Expression constructs were made using two vectors pETM-11 and pOPIN-S (Oxford Protein Production Facility), which is based on the pET28a vector (Novagen). Expression of AKAP79 M was unsuccessful using the pETM-11 construct. The pOPIN-S construct has an N-terminal His tag and a Small Ubiquitin like MOdifier protein (SUMO). Primers were designed that included a specific 15 bp linker sequence homologous to the vector needed for the ligation independent In-Fusion^{TM} cloning (Clonetech). The AKAP79 M was amplified with the primers using the AKAP79 FL as the template. The PCR mix was then incubated with Dpn I to remove any trace template before cleaning with a mini prep kit (Qiagen). The AKAP79 M DNA was then added with the pOPIN-S to the In-Fusion^{TM} mix and heated at 42 °C for 30 minutes before the reaction was stopped with cold TE buffer, and was transformed into XL-1 blue cells. The infusion cloning followed by blue-white screening gave a good mix of white and blue colonies on the agar plate; only white colonies were seen on the positive control. Subsequent PCR screening of the white colonies with the initial primers confirmed the presence of the construct. The colonies were then grown up in 10 ml LB overnight and the plasmid extracted by mini-prep (Qiagen). A sample of plasmid (AKAP79 M Sumo) was sent for Sanger sequencing (Geneservice) using stock T4 forward and reverse primers. The full sequencing confirmed the designed clone (see appendix 3).

4.4.2 AKAP79 M Expression and Purification

The AKAP79 M Sumo construct was transformed into BL21(DE3) cells and expressed at 37°C, before induction at 18°C overnight with 1 mM IPTG. After cell lysis and centrifugation at 48000 g, the supernatant was heated at 65°C for 5 minutes. The sample was centrifuged again at 48000 g for 20 minutes, and supernatant was applied to a Ni^{2+} Affinity column (5 ml HisTrap™ HP, GE), washed with 10 column volumes of His buffer (500 mM NaCl, 50 mM Tris, pH 8.0) and AKAP79M eluted with 250 mM imidazole. After buffer exchange using a HiPrep™ 26/10 (GE) desalting column to remove the imidazole, Sumo protease (0.2 mg) (prepared in the laboratory- see Chapter 3) was added and cleavage allowed to proceed at 30°C for 18 hours. Following cleavage, the protein preparation was reapplied to the Ni^{2+} Affinity column. The flow-through which was cleaved AKAP79 M was collected, desalted once more and further purified using an anion exchange column (5 ml HiTrap™ Q FF, GE. The final protein was analysed by mass spectrometry and confirmed as AKAP79 M.

To prepare uniformly ^{13}C,^{15}N labelled AKAP79 M media containing ^{13}C- labelled glucose and ^{15}N ammonium chloride in a 2xM9 minimal media (MM) was used (See appendix 1 for details of media). The overnight step started with 1 colony grown in 1 ml LB for 4 hours before removal of the LB and resuspension in 40 ml MM overnight at 37°C. The cells were centrifuged at 5000 g for 10 minutes to remove the media before whole cell pellet was used to seed 1 L MM. It was necessary to increase the cell density before seeding the MM or no growth would

occur but this had to be balanced against exposure of the cells to LB; the greater the exposure the greater the decrease in isotopic labelling of the protein.

The expression and purification was performed identically to the unlabelled preparations as described above.

The AKAP79 M was concentrated using an Amicon® Ultra (Millipore, 10K MWCO) to 1 mM. The protein was inserted into an NMR tube and analysed on a Bruker 800 MHz spectrometer at 298 K.

4.4.3 Results

4.4.3.1 Expression and purification

Expression of AKAP79 M from the pETM-11 constructs were screened against all the same bacterial host strains - BL21 RP^{+}(DE3), Rosetta^{TM}(DE3) and B834(DE3) including the TAKARA chaperone plasmids- as previously described with AKAP79 FL. Unfortunately, no bacterial strain gave any detectable expression due possibly to proteolytic cleavage during protein expression. Hence, a new construct was made using the pOPINS vector that includes the genes for an N-terminal His tag and a Small Ubiquitin like MOdifier protein SUMO. The SUMO fusion was originally designed to improve the solubility of expressed proteins. However, in these studies, the SUMO fusion is used to lengthen the protein construct in order to minimise proteolytic cleavage. The fusion protein can be cleaved using a SUMO protease, a highly specific protease which cleaves the fusion protein between SUMO and the desired protein, hence removing the His tag at the same time. The vector is based on the pET28a vector (Novagen).

AKAP79 M Sumo expressed at very high levels. The heat step was introduced to denature the endogeneous bacterial proteins, including proteases; this was found to be effective in improving the long-term stability of the purified proteins. This was a crucial step in the protocol since AKAP79 M, being an unfolded protein, was found to be susceptible to proteolytic cleavage even when only traces of protease were present. The highly specific Sumo protease was used to cleave off the Sumo fusion together with the N-terminus His tag. The optimal working temperature for sumo protease had previously been shown to be in the temperature range 22 – 37°C

although it is has been shown to work at 4°C (Malakhov et al., 2004). In this study, several sets of conditions were tested and it was found that cleavage did not proceed at 4°C. The cleavage was then repeated at 30°C over the time course 1, 2, 4 and 18 hours. As shown in Figure 4.4.1, the 45 kDa AKAP M sumo fusion SDS-PAGE band decreased in intensity to zero after 18 hours confirming the fusion protein cleavage takes a long time. Two gel bands at approximately 30 and 20 kDa increased in intensity during the same period. The 20 kDa gel band is the cleaved Sumo fusion; although it measures 12.4 kDa it has been shown to migrate to approximately 21 kDa (Butt et al., 2005). The 30 kDa band is the cleaved AKAP79 M (18.4 kDa) that migrates to a greater molecular weight that is attributed to lack of tertiary structure.

For preparation of ^{13}C, ^{15}N AKAP79 M, it was necessary to increase the cell density before seeding the MM; otherwise no growth would occur. However, it is necessary to minimised isotopic dilation by the LB media; only a minimal quantity of the LB culture was used for the seeding. The purification of ^{13}C, ^{15}N AKAP79 M was performed identically to the unlabelled preparations (section 4.4.1)

The purification of unlabelled AKAP79 M yield approximately 20 – 30 mg/L where as the double labelled preparation gave a lower yield of approximately 10 – 15 mg/L

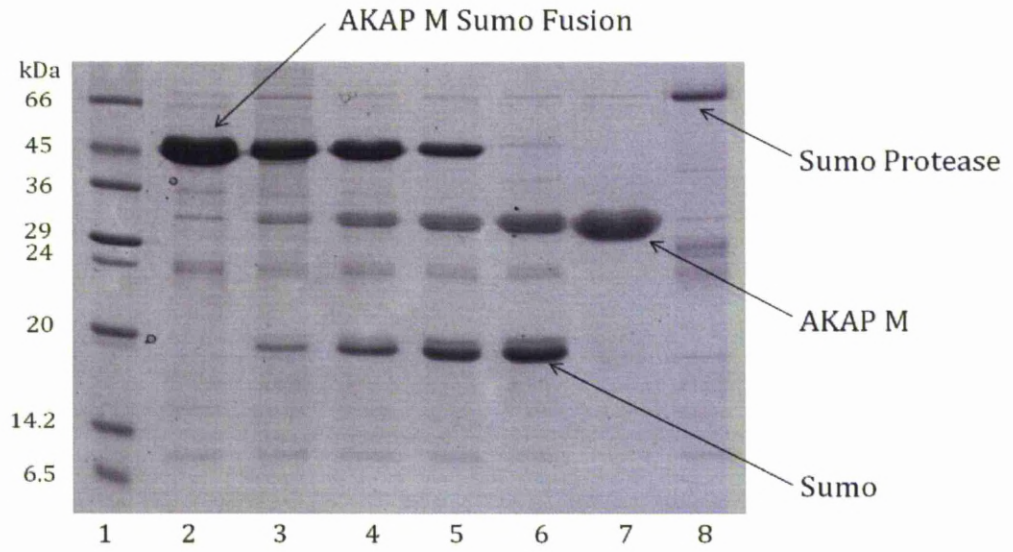


Figure 4.4.1. Effect of time on AKAP M Sumo fusion cleavage analysed by 15 % SDS PAGE. AKAP79 M was first purified using Ni^{2+} affinity chromatography (5 mL HisTrap HP, GE) and buffer exchanged to remove imidazole. To the eluant an aliquot of Sumo protease (0.2 mg) was added to give a T=0, Lane 2, then incubated at 30 °C with gentle agitation. Samples were taken at T= 1, 2, 4 and 18 hours, lanes 3 to 6 respectively. After 18 hours the sample was passed back through a Ni^{2+} affinity column to separate the cleaved AKAP79 M from the sumo fusion containing the N-terminal His Tag and the sumo protease, Lane 7. Sumo protease was run in lane 8.

4.4.3.2 NMR Spectroscopy

The AKAP79 M was concentrated using an Amicon® Ultra (Millipore, 10K mwco) to 1 mM. The protein was inserted into an NMR tube and analysed on a Bruker 800 MHz spectrometer at 298 K.

The $^{15}N^{1}H$ back bone resonance assignments were completed by Dr Marie Phelan, using triple resonance CBCANNH and CBCA(CO)NNH spectra. The CBCANNH strongly identifies the Cα and Cβ chemical shifts of its own residue correlated to the corresponding NH group and weakly identifies the preceding Cα and Cβ. Whereas CBCA(CO)NNH only correlates the NH group to the preceding Cα and Cβ chemical shifts. Using both spectra together is was possible to walk backwards

following the chemical shifts of each of the preceding residues in the sequence; this is shown graphically in figure 4.4.2 (Protein NMR - A Practical Guide by Dr Victoria Higman). Assigned AKAP79 M is shown in figure 4.4.3.

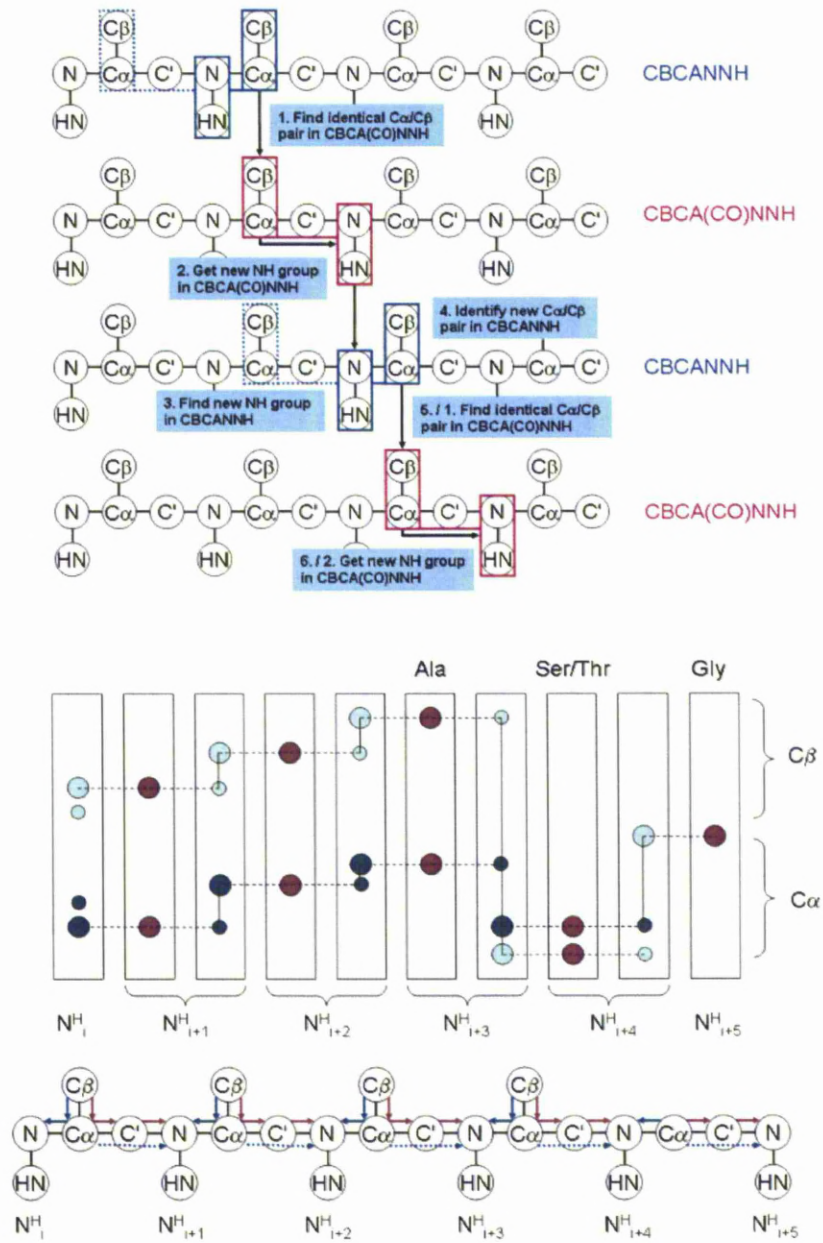


Figure 4.4.2. Graphical backbone assignment. Top, showing NH groups linked to the next NH group. Middle shows a representation of the assignment as seen from the software. With the corresponding NH linking, bottom where Cα dark blue and Cβ light blue. Taken from Protein NMR - A Practical Guide by Dr Victoria Higman

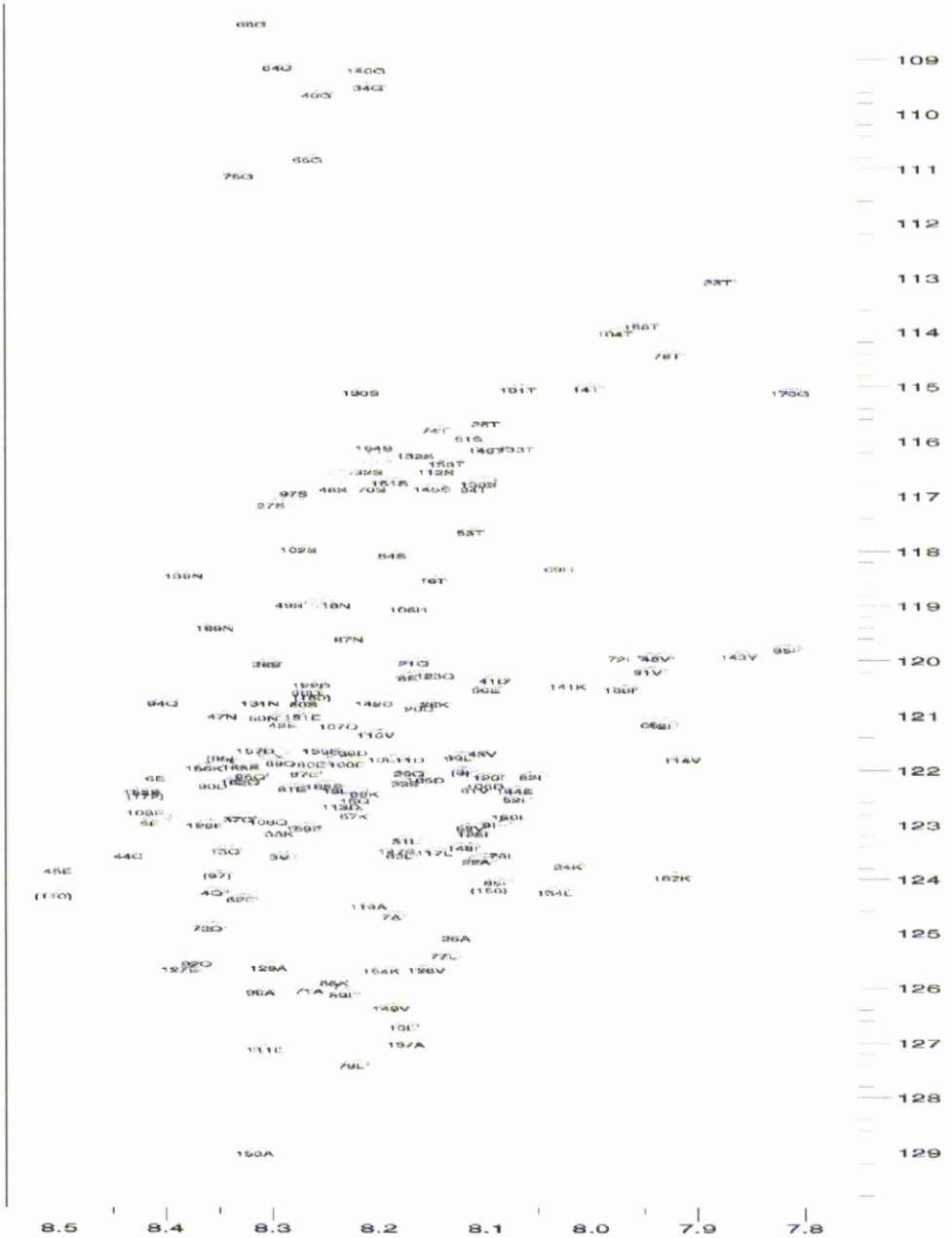
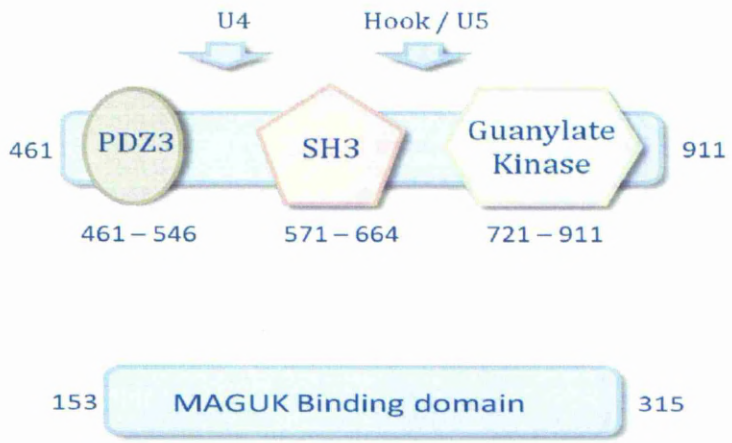


Figure 4.4.3. ^{15}N-^{1}H HSQC spectrum of ^{13}C/^{15}N AKAP79 M. The chemical shift dispersion is narrow, reminiscent of a predominantly unstructured protein. The $^{13}C^{15}N$ ^{1}H backbone and Cβ resonances have been completely assigned using triple resonance CBCANNH and CBCA(CO)NNH spectra.

4.5 AKAP79 M NMR Binding Studies

AKAP79 had been reported to bind to PSD95 and SAP97 via a large region corresponding to AKAP79 M (aa153-315, 18.4 kDa) (Robertson et al., 2009). Previously, AKAP79 was shown to interact with the SH3 and GK domains of PSD95 (Colledge et al., 2000). To this end the SAP97 construct Δ461 (53 kDa) that contained a PDZ domain coupled with the SH3 and GK domains was investigated for binding to AKAP79 M.



Using the NMR method to investigate this interaction, unlabelled SAP97 Δ461 was added and mixed to form a complex. The resultant ^{15}N-^{1}H HSQC spectrum is shown in Figure 4.5.1 and a chemical shift histogram in Figure 4.5.2. To express changes in the chemical shifts of the individual amide pairs a compound chemical shift change (in ppm) was defined as $\Delta\delta = (\Delta^{1}H)^{2}+0.15\ (\Delta^{15}N)^{2}$, where ΔH and ΔN are the chemical shift differences between the apo and bound in the proton / nitrogen dimension respectively. The weight to the amide nitrogen and proton

chemical shifts measurements was accorded to their relative measurement error. This differs from one experiment to the next, and should be estimated on a case-by-case basis, 0.15 was used in this case.

No significant shift changes were observed for the resonance of AKAP79 M, suggesting that there was no interaction between the two proteins. The slight movements of a few resonances, the greatest being 0.077 ppm, are likely caused by small changes in pH between reference AKAP79 M and the complex.

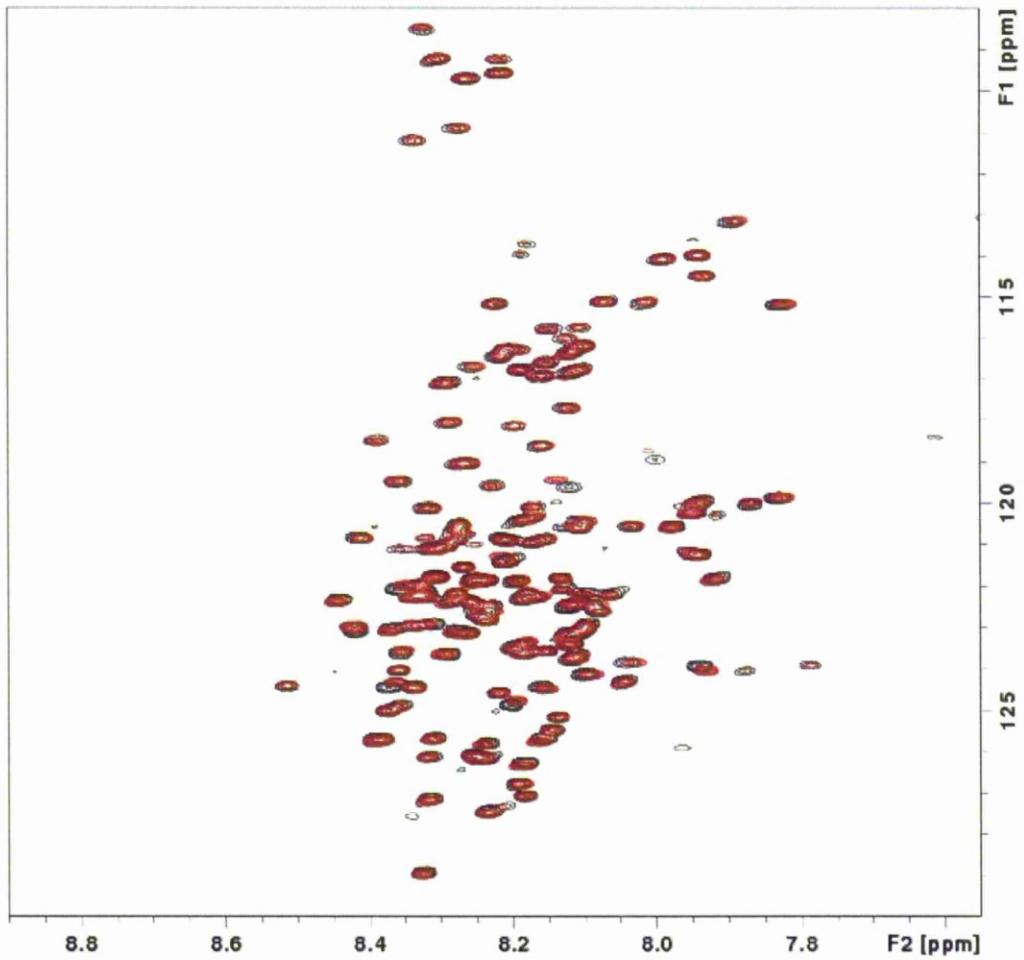


Figure 4.5.1. ^{15}N-^{1}H HSQC spectrum of 13C/15N AKAP79 M in complex with SAP97 Δ461. Black = free protein; Red = complexed protein. Negligible shift movement was observed when unlabelled SAP97 Δ461 was added at ratio 1:2.

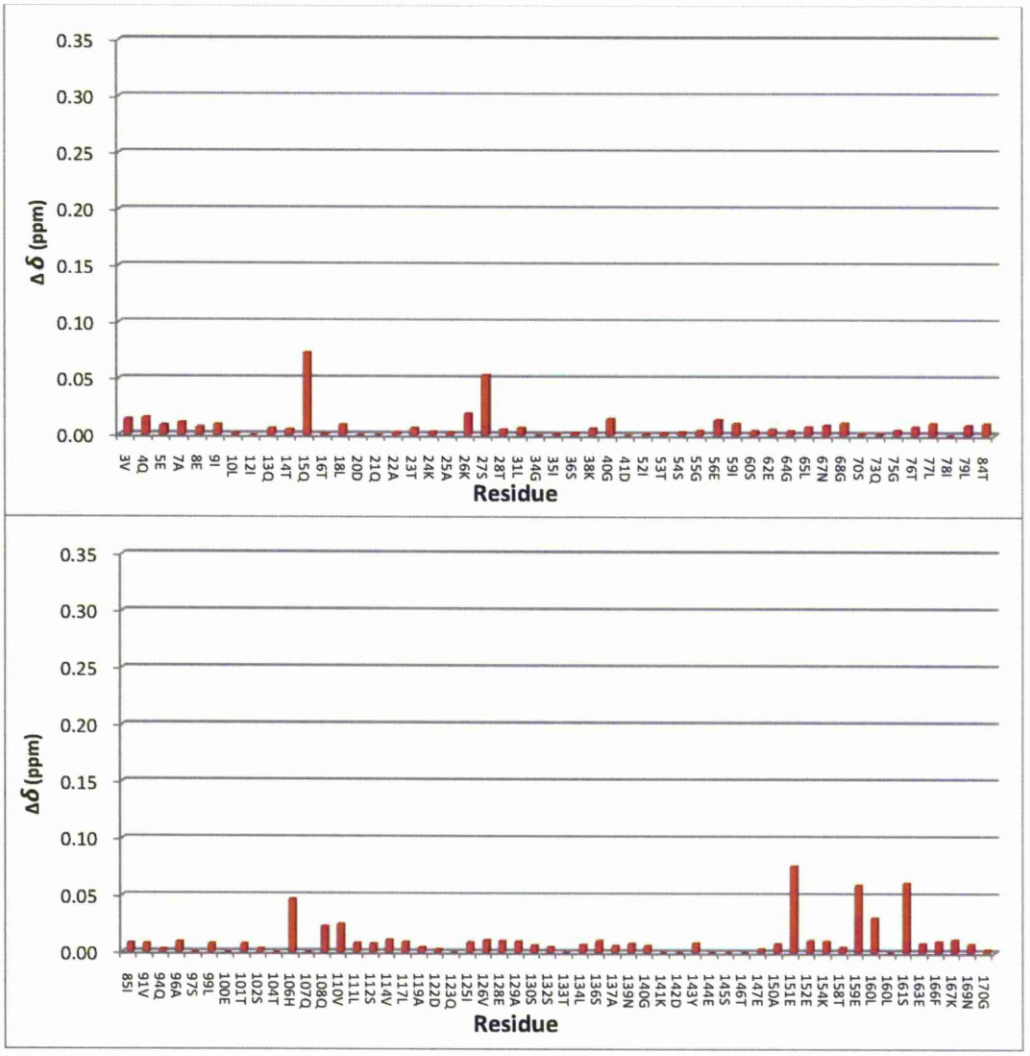


Figure 4.5.2. Bar chart of chemical shift from AKAP79 M in complex with SAP97 Δ461. Chemical shifts were measured between AKAP79 M and the 1:2 SAP97 Δ461 2D ^{15}N HSQCs respectively. Chemical shift change derived as $\Delta\delta = (\Delta^1H)^2+0.15\ (\Delta^{15}N)^2$ where ΔH and ΔN are the chemical shift differences between the apo and bound in the proton / nitrogen dimension respectively.

4.6 AKAP79 M Binding to SAP97 by ITC

To further probe the binding of AKAP79 M and SAP97 Δ461, isothermal calorimetry (ITC) was used. This technique measures the tiny changes in enthalpy that occur during binding of proteins and ligands. The highly concentrated ligand is titrated into the protein and the heat exchanged recorded. This technique is useful because it requires no labelling or tagging of proteins and uses very small volumes, approximately 70 µl Ligand and 300 µl protein in the cell, although, it often requires high concentrations.

Unlabelled 2 mM AKAP79 M was titrated into 0.2 mM SAP97 Δ461, but showed no binding (Figure 4.6.1). Protein and ligand had been dialysed overnight to ensure no buffer mismatches and then concentrations measured using nano drop. Several concentrations of ligand to protein were tried with no success on different days and with two preparations of the proteins in question.

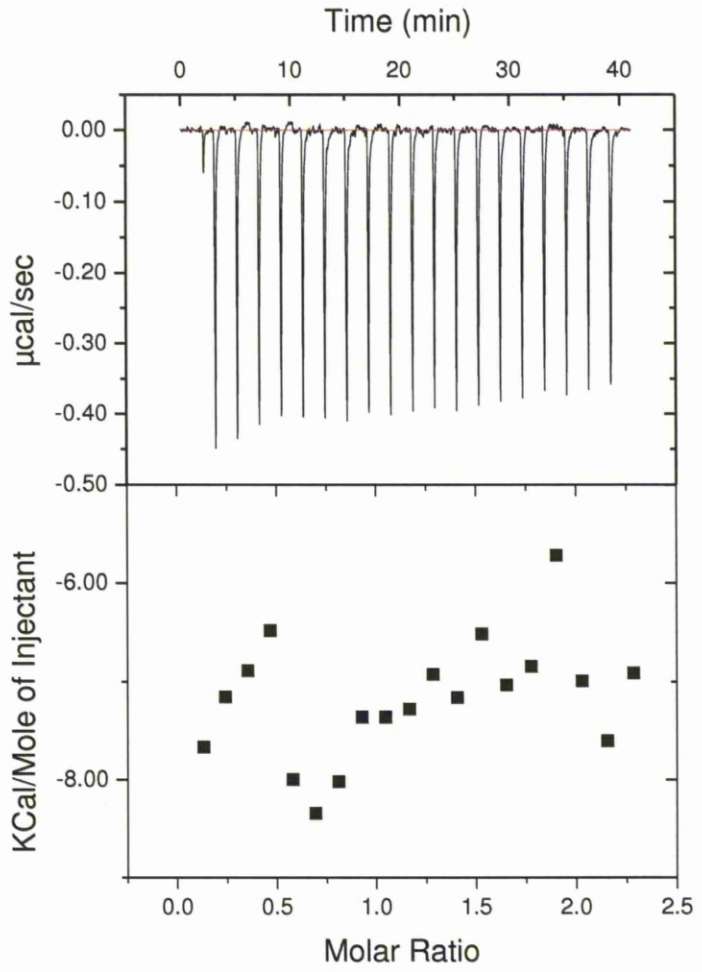


Figure 4.6.1. Isothermal Calorimetry (ITC) of AKAP79 M with SAP97 Δ461. 2 mM AKAP79 M was titrated into 0.2 mM SAP97 Δ461 at 25°C in 50 mM Tris HCl, 50 mM NaCl pH7.5. The thermogram, top and integrated titration, bottom showed no binding.

4.7 Discussion

AKAP79 is largely unfolded as was shown by the 1D NMR. This is most likely the reason for the poor expression of the full length and shorter domains since unstructured proteins are more susceptible to in-cell proteolysis. The use of chaperones and SUMO fusions alleviated the proteolysis problems and enabled adequate quantities of *E.Coli* expressed recombinant AKAP79 and its domains to be purified for structural characterisation.

The use of the SUMO tag was shown as essential to express the longer AKAP79 M domain. It is possible that other fusion proteins may have also had the same effect such as glutathione-S-transferase (GST) or maltose binding protein (MBP) but these were discounted as they are often harder to purify and therefore give rise to lower yields when compared to His tag purification. Added to this is the cleavage of the SUMO tag uses SUMO protease a highly specific enzyme that recognises a specific α-Helix rather than just a specific sequence. This enables the cleavage site to be flush with the expressed protein and thus leaves no overhanging residues on the protein. In further work the other domains, AKAP79 C / N will be subcloned into SUMO vectors to enable there expression and purification.

AKAP79 is a known scaffold protein which is able to bind several proteins at the same time. These target proteins have been reported to include PKA, PKC, CaM, PSD95 and SAP97. The binding sites of these proteins have been mapped out on the primary sequence of AKAP79. The region that interacts with PKA and PKC adopt a helical structure when in complex with the two kinases. Likewise for the

CaM binding region (see Chapter 5). The fact that AKAP79 is unfolded may actually be necessary for its function. With no steric hindrances the protein is able to bind and possibly coil around its many binding partners or to bind many of the binding partners simultaneously when possible.

In the present studies, the binding between AKAP79 was investigated further since it had been widely reported in the literature. In a related study in the laboratory using NMR, only extremely weak interactions were detected between AKAP79 FL and SAP97 Δ461. AKAP79 M was identified by Colledge et al 2000 as binding site for both SAP97 and PSD95. However, no interactions were detected between AKAP79 M and SAP97 Δ461 using either ITC or NMR. There could be several reasons for this. It is likely that AKAP79 and SAP97 on their own do not bind to each other but require additional components, secondly, that the methods used to identify these interactions are cellular based techniques which can sometimes provide false positives or lastly, that the methods used were unable to identify the binding. Taking each point in turn; firstly AKAP79 does not bind to SAP97 although the binding was proposed by Colledge et al 2000. Their data was based on GST pull downs. Recombinant SAP97 was bound to the resin and AKAP79 from brain extract was passed over the beads, washed and eluted with SDS before western blotting identified that the AKAP79 had bound. They did note that the amount of AKAP79 bound to SAP97 was lower than that of PSD95. That possibly suggests a different mechanism of binding or possible non-specific binding of AKAP79. Another problem with the pull down was that the AKAP79 was from brain extract that may have already been complexed with necessary binding partners and was only checked for purity through western blotting. It is also

conceivable that the AKAP79 was already bound to PSD95 from the brain extract. After probing the pull down with anti-AKAP79 antibody the blot was stripped and re-probed with anti-GST antibody, many bands were seen in the SAP97 lane as seen in Colledge et al 2000. The bands shown may have been degradation products of SAP97; if they are they may lead to increased non specific binding. It has been observed that AKAP79 will bind strongly to other proteins, in particular heat shock proteins that could only be removed through rigorous purification (see AKAP79 FL purification above).

The second possible reason for no binding is that the AKAP79 M sequence does not contain the binding site to SAP97. It has been suggested that binding of AKAP79 to calcineurin inhibits the binding of AKAP79 to PSD95 (Colledge et al., 2000), that is calcineurin binds competitively with PSD95. The calcineurin binding site is at residues 318 – 357 on AKAP79 (Dell'Acqua et al., 2002) and thus not within the AKAP79 M sequence (153 – 315). This would suggest that the SAP97 binding site was not located in AKAP79 M. However, Dell'Acqua 2006 believes the contrary; that both MAGUKS and CaN can bind to AKAP79 at the same time and that it is the binding of PKA that inhibits the CaN interactions. More experiments are needed to resolve this controversy.

The third possible reason for the absence of binding between AKAP79 M and SAP97 is the binding site on SAP97 is not present or masked. PSD95 was shown to bind to AKAP79 within the region of the SH3 – GK domain (Colledge et al., 2000). There is one major difference between PDS95 and SAP97 in this region; SAP97 possesses an extended Hook region that may block the binding site unless another binding partner to SAP97 can lead to a displacement of the Hook region. It may be

that SAP97 full length is required with the intramolecular interactions being necessary to afford a conformation that enables AKAP79 to bind.

For future work, the AKAP79 Full length, C and N domains could be cloned into the pOPINS vector, expressed and purified before repeating the binding studies. Full-length SAP97 full length could also be expressed and purified to test whether intramolecular forces are required for the binding to occur.

CHAPTER 5

AKAP79 INTERACTIONS WITH CALMODULIN

5.0 Introduction

The AKAP79N domain, 1 – 153 residues contains three highly basic binding regions, 31 – 52, 76 – 101 and 116 – 145. The first domain, 31 – 52 contains a hydrophilic helix that had been shown to bind to PKC as well as membranes (Klauck et al., 1996). Originally it had been purported that the second basic region 76 – 101 may bind Calmodulin, CaM (Carr et al., 1992); however, it was subsequently shown that the polypeptide region 31-52 interacted with CaM. This competition between CaM and PKC for the same region on AKAP79 forms the basis for competitive release of PKC by CaM (Faux and Scott, 1997). CaM has also been reported to regulate the binding of AKAP79 to membranes both *in vitro* and *in vivo*, supporting the hypothesis that AKAP79 may have multiple binding sites for CaM (Dell'Acqua et al., 1998).

The CaM binding motifs are defined by the position of the hydrophobic residues. The target sequence AKAP79 31 – 52 gave a binding motif 1-11, this did not follow any of the regular motifs such as 1-10 and 1-14. However, a recent structure of the CaM:calcineurin A complex, previously listed in the calmodulin target database to bind through a 1-5-8-14 motif was actually shown to have a 1-11 binding mode (Majava et al. 2009)

This chapter aims to characterise at the interaction between AKAP79 and CaM. It attempts to ascertain the positions on CaM were AKAP79 binds using both a short synthetic peptide encompassing residues 31 – 52 and longer recombinant peptides.

5.1 NMR Studies Protein – Peptide interaction

To understand the structural basis of the binding of CaM to AKAP79 the interactions between a synthetic $AKAP79_{31-52}$ peptide and a double labelled CaM were studied by NMR.

5.1.1 Calmodulin NMR Spectroscopy

A $^{13}C/^{15}N$ rat brain calmodulin sample was expressed and purified (see Chapter 3). A 0.2 mM sample (10 mM CaCl2, 50 mM Tris, pH 6.5) was used for the NMR experiments and analysed on a Bruker 600 MHz spectrometer at 298 K. The acquired $^{15}N/^{1}H$ HSQC was assigned predominantly by transferring the assignment of crane toad CaM (BMRB Database Entry 6541); both CaM shared significant sequence homology. Peaks that could not be assigned in this manner (due either to different buffer and temperature conditions or differences in the amino acid sequence), were assigned from a new set of triple-resonance data collected; Dr. Marie Phelan from the laboratory performed these resonance assignments. The resonance assignment for rat CaM was complete apart from 5 residues (first three at the N-terminus, H107 and N111) (Figure 5.1).

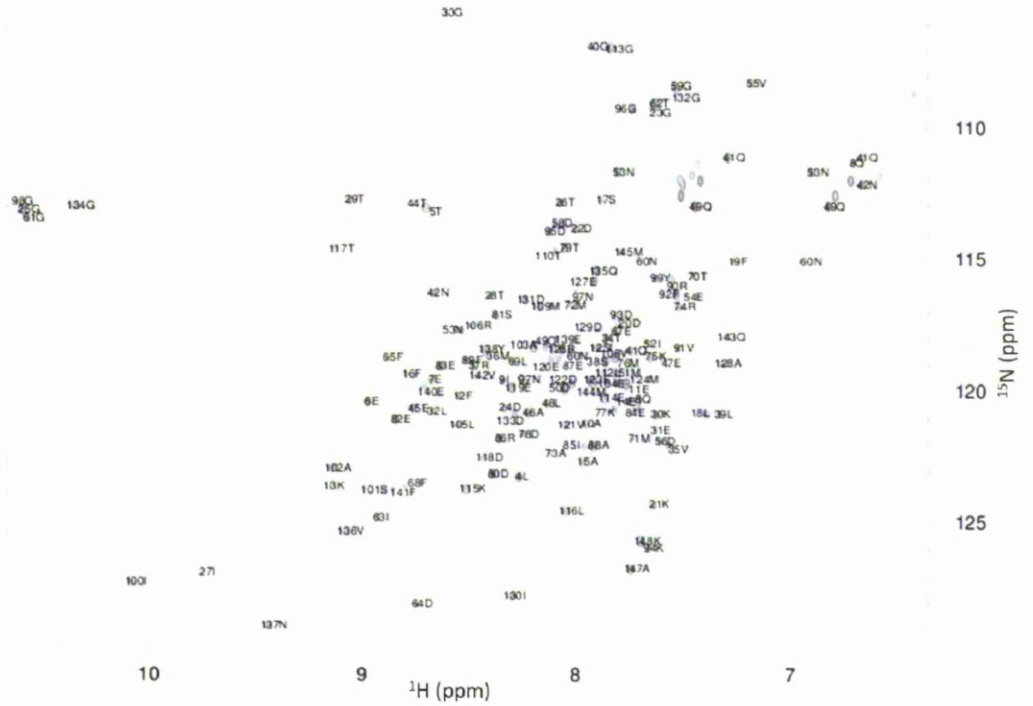


Figure 5.1. Assigned 2D ^{15}N HSQC spectra of calmodulin. 0.2 mM CaM in 10 mM $CaCl_2$, 50 mM Tris, pH 6.5 at 300 K.

5.1.2 CaM - AKAP79$_{31-52}$ Titration

Unlabelled synthetic AKAP79$_{31-52}$ was purchased lyophilised. Recombinant $^{13}C/^{15}N$ labelled CaM was lyophilised from water. Titration of the peptide was performed in the following manner. Two samples are first made in identical buffers: uncomplexed CaM (sample A) and CaM:peptide complex in a concentration ratio of 1:10 (sample B). The protein concentrations are identical in both samples. The samples were made up in 10 mM $CaCl_2$, 50 mM Tris at pH6.5. The samples of the two initial solutions were checked to ensure that the pH's were identical and adjustments were made if necessary. 2D ^{15}N HSQCs and ^{13}C HSQC were acquired for each of these starting samples and the spectra represented the start and end-

points of the titration series. Equal aliquots from samples A and B were and store in separate tubes. The withdrawn aliquot (50 μl) from A was added to B and the aliquot from B added to A. After thorough mixing, the ^{15}N-^{1}H HSQC and ^{13}C-1H spectra are acquired for the new samples A and B. After this, another set of equal aliquots were again removed, and each added to other tube and the HSQC spectra. The idea behind this method of doing a 'titration' is to ensure (i) the buffer conditions remain identical throughout the titration with no dilution errors; (ii) the protein concentration remains the same for all titration points. In all, data for 11 titration points were acquired. The 2D ^{15}N HSQC data was collated and overlaid (Figure 5.2).

With increasing peptide concentration in the sample, the chemical shifts of many of the CaM resonances changed gradually. It was also evident that the chemical shifts were in different directions suggesting the shifts were caused by binding between the peptide and CaM. If shift changes were caused by changes in sample conditions, such as salt, then these changes would have been systematic, that is non-specific shifts and all in the same direction. In addition, line-broadening was also observed, in particular at the higher peptide concentrations.

The shift differences in the 28 most resolved residues were measured at each titration point, and were used to calculate the K_D values for the interaction between CaM and $AKAP79_{31-52}$ peptide. Table 5.1 shows the 8 residues that gave the greatest binding, With K_D between 0.4 and 1 mM. The NMR data suggests very weak binding of CaM for $AKAP79_{31-52}$ peptide. A selection of the binding curves and the corresponding chemical shifts are shown in Figure 5.3. Unfortunately, the binding curves do not reach saturation and any further increase in the concentration

of the peptide would have been impractical due to insolubility issues with the peptide.

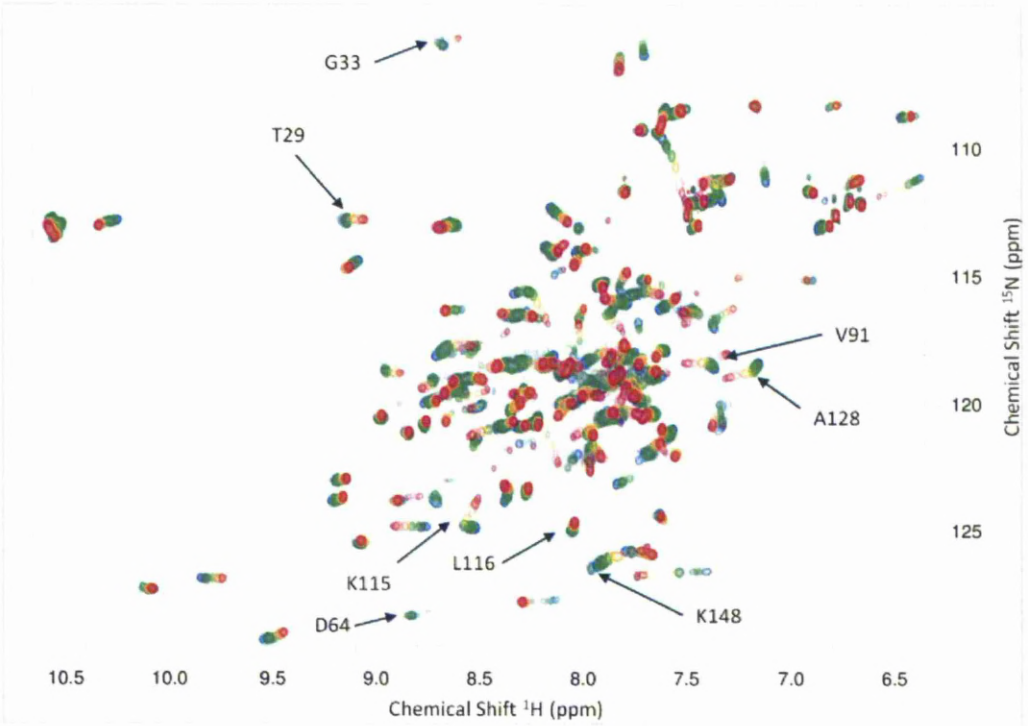


Figure 5.2. Superimposed 2D ^{15}N HSQC spectra of calmodulin with additions of AKAP79$_{31-52}$. Ratios of 1:0.5, 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, and 1:10 CaM to peptide were used. The colour code for the residues moves from magenta for CaM without peptide through the rainbow to violet for 1:10 CaM:peptide. The shifts from a few selected residues were analysed in detail.

Table 5.1. Chemical shift differences and derived binding constant of calmodulin interactions with AKAP79$_{31-52}$ obtained from NMR titration. Only selected residues where the data was clear are shown. K_D derived using equation $[PL]=[L_{Tot}]+[P_{Tot}]+KD\pm\sqrt{([L_{Tot}]+[P_{Tot}]+K_D)^2-4[L_{Tot}][P_{Tot}]}$ Where P_{Tot} is total protein and L_{Tot} is total ligand

Assign F1	Traj Dist	Shift Dist (Δδppm)	Fit Error	Num Peaks	K_D (M)	K_D error
K115	0.228	0.173	0.101	11	3.93E-04	7.25E-05
T29	0.105	0.091	0.047	11	4.26E-04	8.11E-05
V91	0.181	0.150	0.051	11	6.07E-04	7.59E-05
K148	0.297	0.258	0.151	12	6.67E-04	1.43E-04
A128	0.219	0.178	0.049	12	6.82E-04	6.68E-05
L116	0.073	0.064	0.017	11	7.19E-04	7.57E-05
D64	0.197	0.089	0.106	12	7.97E-04	1.88E-04
G33	0.163	0.090	0.058	11	9.99E-04	1.81E-04

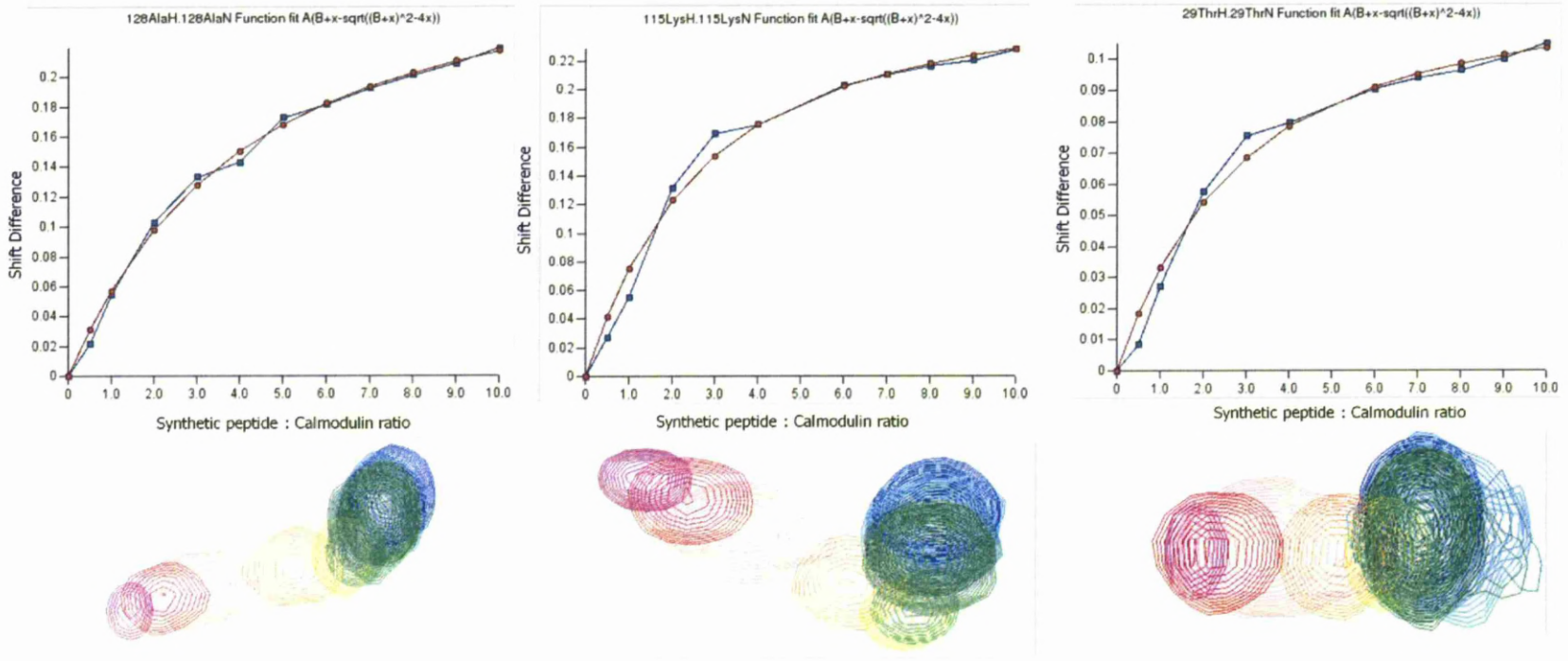
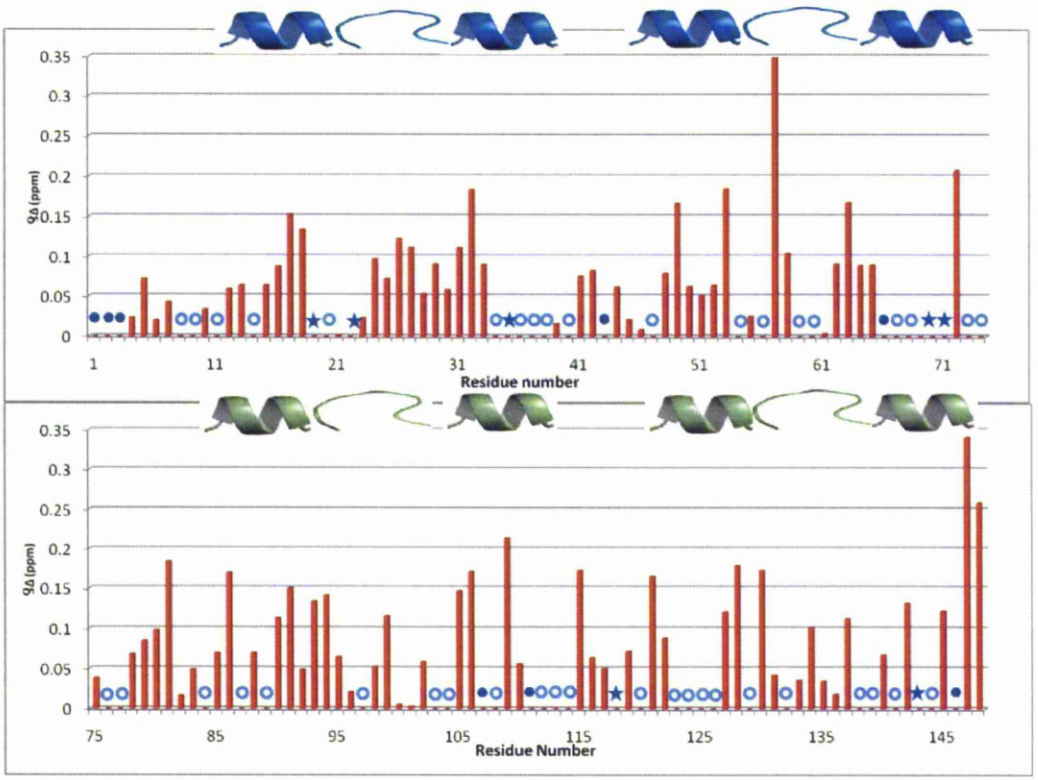


Figure 5.3. **Chemical shift differences derived binding constant from calmodulin : AKAP79$_{31-52}$ titration.** Curves show actual measured shift differences (blue) and the regression curve from which the K_D is derived from (red). Below each curve the actual chemical shifts from each titration of CaM:AKAP79$_{31-52}$ peptide Ratios of 1:0 Magenta, 1:0.5 red, 1:1 pink, 1:2 orange, 1:3 yellow, 1:4 green, 1:5 lime, 1:6 forest, 1:7 marine, 1:8 cyan, 1:9 blue, and 1:10 navy.



1 ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMAR 74

75 KMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK 148

Figure 5.4. Bar chart of chemical shift as a function of residue number. Chemical shifts were measured between the CaM peptide free and the 1:10 CaM:peptide 2D ^{15}N HSQCs respectfully. Enclosed circles show unassigned peaks, open circles show merged peaks in which the chemical shifts could not be accurately attributed to the residues and stars show residues of peak broadening. $\Delta\delta = (\Delta^{1}H)^{2}+0.15\ (\Delta^{15}N)^{2}$. The pictorial above the bar chart shows the approximate position of the helix-loop-helix motif of each of the EF-hands. CaM sequence underneath graphs, red shows residues that shifted greater than 0.15 ppm and in purple residues that achieved strong binding affinities. The blue line above the sequence shows N-Lobe, cyan linker region and green C-terminal lobe.

Table 5.2. Distribution of calmodulin residues affected by chemical shift differences.

N-terminal Lobe (1-69)	Linker region (70-85)	C-terminal Lobe (86-148)
S17	M72	R86
T29	S81	V91
L32		R106
G33		M109
Q49		K115
N53		L116
A57		V121
I63		A128
D64		I130
		A147
		K148

Figure 5.4 is a bar chart showing the magnitude of chemical shift changes. Unfortunately, a large proportion of peaks merged with other peaks thereby disguising the end points of the chemical shift; these were mainly found in the centre of the HSQC, which is the most crowded region. A collection of residues also disappeared due to line broadening with increasing peptide concentrations.

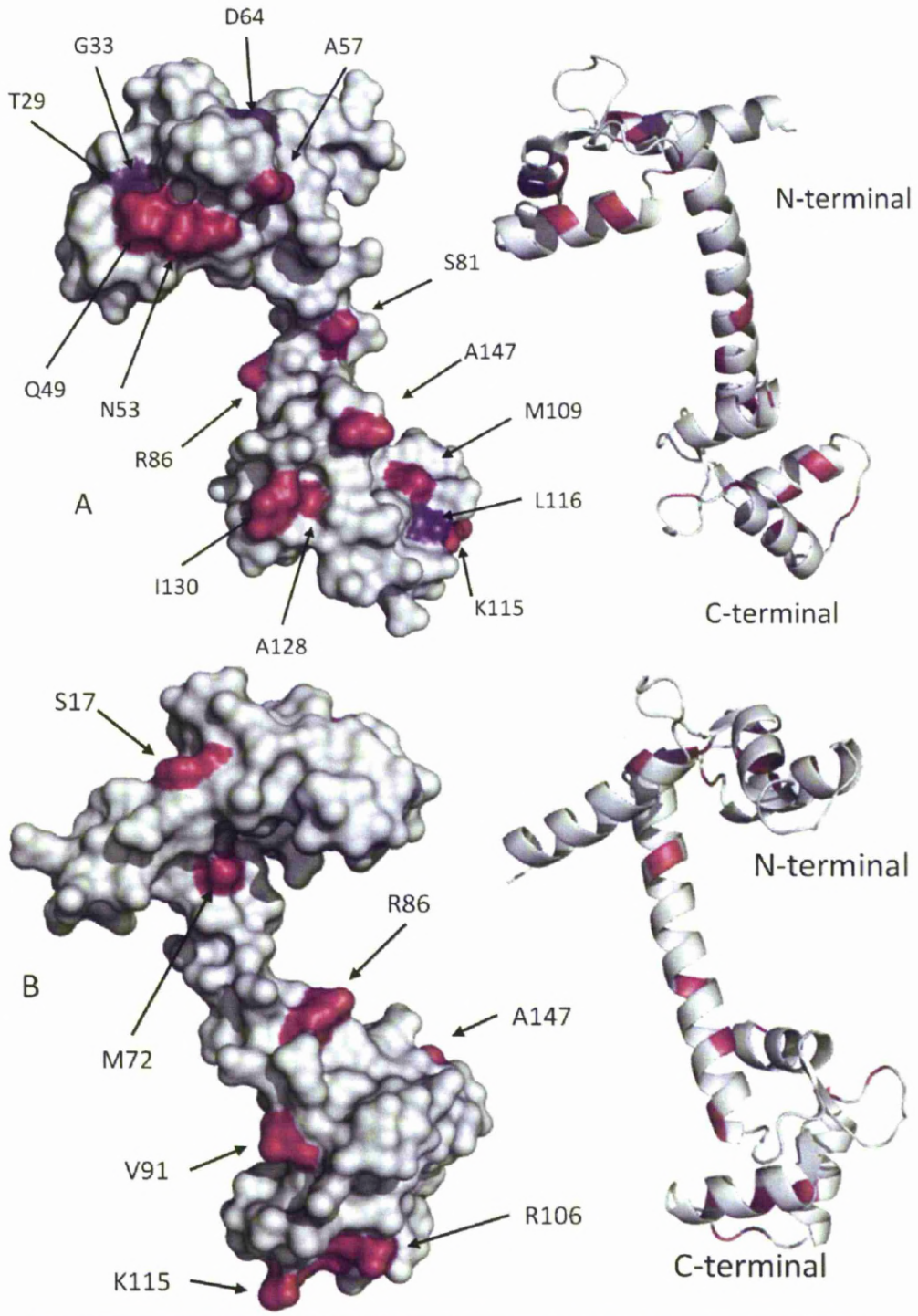
The chemical shift change (in ppm) was defined as $Y= (\Delta^{1}H)^{2}+0.15 (\Delta^{15}N)^{2}$, where the weight to the amide nitrogen and proton chemical shifts measurements was estimated to be 0.15.

Residues with an overall chemical shift of 0.15 ppm were classified as strongly affected by the peptide binding, this gave 18 residues. A further four residues were also classed as strongly affected due to their stronger binding affinities achieved (marked in purple on sequence below histogram). The affected residues were also tabulated to show distribution across CaM domains, Table 5.2.

The strongly affected residues based on the interpreted chemical shift changes were then mapped onto the open form, human CaM crystal structure (pdb 1CLL)

(Chattopadhyaya et al., 1992), that is homologous to the rat brain structure, Figure 5.5. The mapping clearly shows regions on the N-terminal lobe that are affected by the presence of $AKAP79_{31-52}$ peptide. More specifically, residues from two of the α-helices and a β-sheet appear to be involved in interactions with the peptide. It is interesting to note that these helices and β-sheet fold to form a channel; it is possible that the peptide binds in this channel. There are further areas of interaction along the central α-helical span and then onto a further area of interaction at the C-terminal lobe. The interactions shown on the 180° image shows fewer possible areas of interaction.

In CaM peptide complexes the residues that are most readily perturbed, providing a hydrophobic binding surface are the methionine residues (Vogel et al., 2011). Only M_{72} was shown to be perturbed, unfortunately, the peak shifts of the other methionine residues were unable to be measured due to the peaks merging and in some cases through line broadening that is indicative of fast-intermediate chemical exchange between free and bound forms of the peptide. However, with further analysis it may be possible to resolve these peaks and peak shifts.



1 ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMAR 74

75 KMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK 148

Figure 5.5 See next page for legend.

Figure 5.5.Chemical shift map of interactions between CaM and AKAP79$_{31-52}$ peptide. A surface representation of CaM with the residues affected by interactions with AKAP79$_{31-52}$ peptide depicted. Residues coloured magenta show interactions from chemical shifts >0.15 ppm .A ribbon model is next to each surface representation. B shows CaM rotated through 180°. Blue line above shows N-Lobe, cyan linker region and green C-terminal lobe.

From the interaction map the one possible orientation of the peptide would be if it spanned both lobes, binding to both lobes and the α-helical linker simultaneously. However, the AKAP79$_{31-52}$ peptide is only 22 residues long, therefore too short to span the gap, Figure 5.6. Since there is substantial line-broadening in the NMR spectrum of the CaM- the AKAP79$_{31-52}$ peptide complex, it was difficult to determine the full structure of the complex by NMR. Further work is underway in the laboratory using other structural techniques.

Two modes of the known CaM-peptide complexes can also possibly explain the current NMR data: (i) the classical compact CaM-peptide structure where the N and C- domains sandwich the peptide with the collapse of the helical linker (a 1:1 stoichiometry), and (ii) the new X-shape dimer of CaM binding to two molecules of peptide (a 2:2 complex) These two modes are represented in Figure 5.6 (de Diego et al., 2010; Majava and Kursula, 2009). Stoichiometry data would be required to verify which mode of binding is most appropriate for the CaM- AKAP79$_{31-52}$ peptide complex.

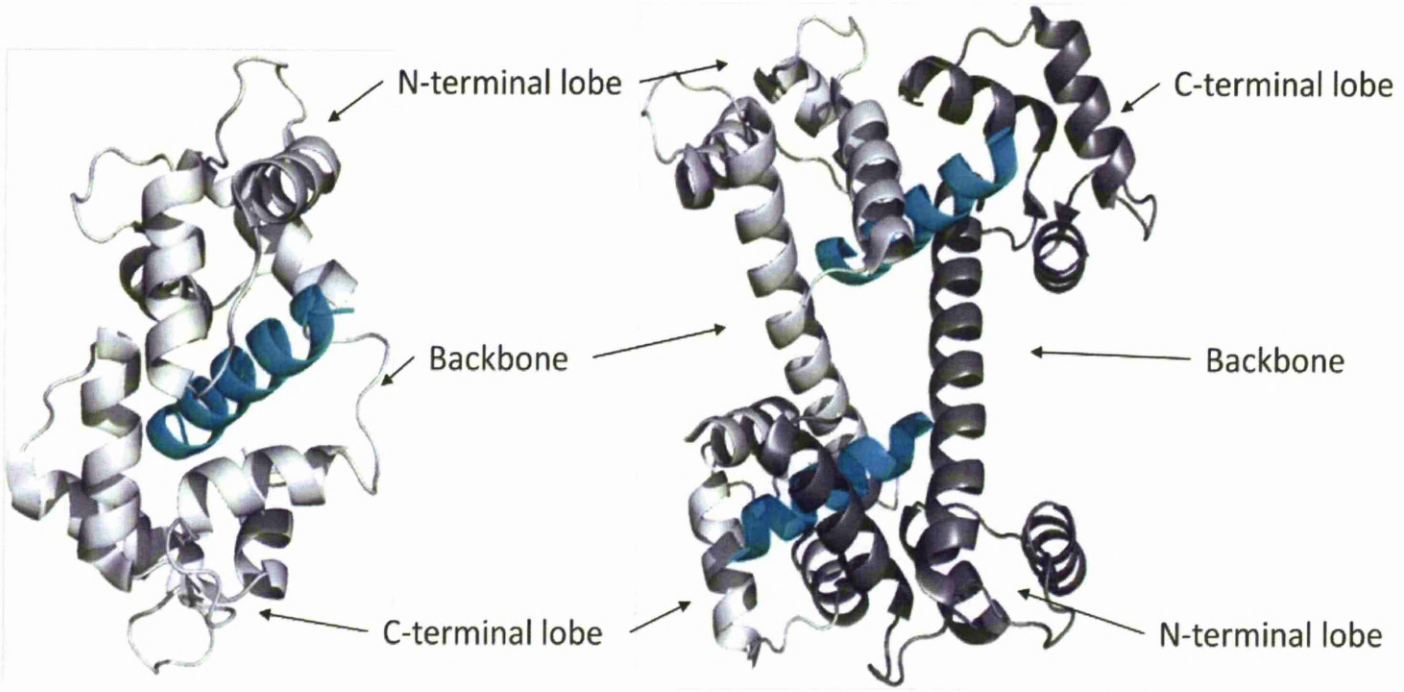


Figure 5.6. Ribbon diagrams of interactions between CaM and DAPK peptide and CaN peptide. Binding structures show 1:1 stochiometry collapsed lobe binding, left (de Diego et al., 2010) and the 2:2 stoichiometry X-shape CaM dimer binding, right (Majava and Kursula, 2009).

Using the compact model from as a template, pdb 1YR5 (de Diego et al., 2010) the residues affected by the binding of $AKAP79_{31-52}$ peptide were mapped and depicted in stick form, Figure 5.7. The model is much improved over the open CaM model. The C-terminal lobe residues are mostly positioned close to the peptide helix. However, the N-terminal lobe does not fit the model as well, with many of the residues far from the peptide surface. This indicates that a further movement in the N-terminal lobe will be required to reposition the perturbed residues in close proximity to the peptide. Further biophysical analysis is required to prove the proposed complex.

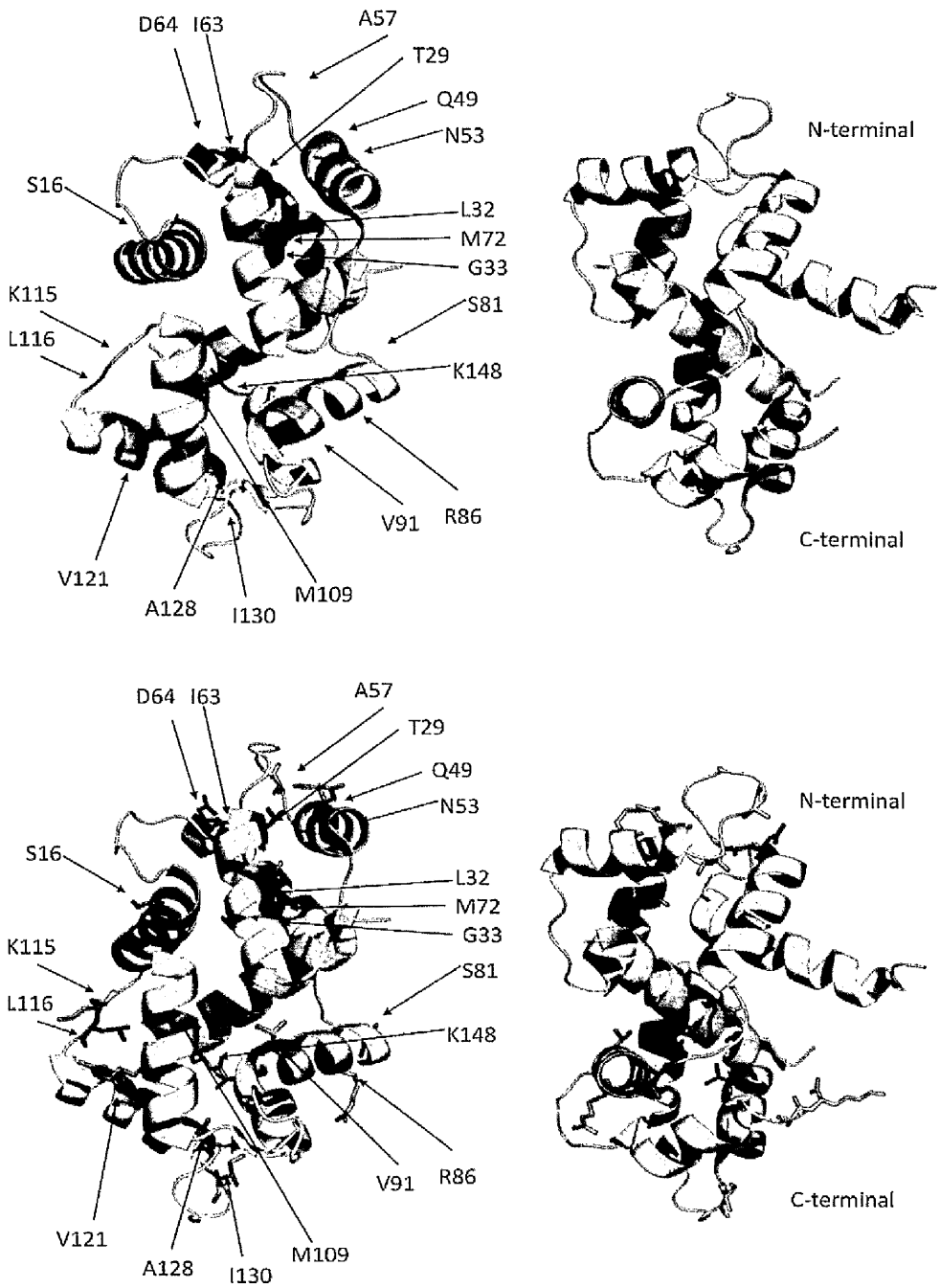


Figure 5.7. Chemical shift map of interactions between CaM and model of AKAP79$_{31-52}$ peptide based on interactions between CaM and DAPK peptide. A simple AKAP79$_{31-52}$ peptide model showing chemical shifts was created using Pymol assuming that binding structures show 1:1 stochiometry and using CaM collapsed lobe model 1YR5 (de Diego et al., 2010). A ribbon model of CaM with the residues affected by interactions with AKAP79$_{31-52}$ peptide, below, depicted in stick form. Residues coloured magenta show interactions from chemical shifts >0.15 ppm, the right shows CaM rotated through 180°.

5.2 Recombinant AKAP79 Calmodulin Binding Domain

To investigate the binding of CaM to AKAP79 further it would be necessary to repeat the NMR binding experiments with $^{13}C/^{15}N$-labelled in complex with unlabelled CaM. The cost of a double labelled synthetic peptide was prohibitive so a recombinant peptide was made. The peptide was also made longer for ease in cloning and purification plus it was thought the larger peptide might have increased the binding affinities.

5.2.1 Recombinant AKAP79 CBD Sequence Identification

The highly basic region, 31 – 52 was predicted to be α-helical, using the computer programme, AGADIR (Munoz and Serrano, 1997), which is a prediction algorithm for α-helix and coils based on transition theory. AGADIR identifies each residue and calculates percentage helical content for that residue. The 31-52 sequence showed a helix would be present but was weak, between 0.1 and 3.8% (Table 5.3). To test whether longer peptide sequences still had the propensity to form α-helices or whether it was possible to improve the α-helices the short 31 – 52 peptide sequence was elongated towards both the N and C-termini and re-ran through the AGADIR program.

When the starting residue was moved to Ala_{24} the helix strength increased from 3.8 to 5.2%. When the residue was moved even further to Ala_{19} the helix strength increased further to 7.0%. Increasing the length towards the C-terminus did not improve the helical propensity, although a second weak helix was found.

To understand and characterise the AKAP79 Calmodulin binding domain further three elongated constructs were cloned, AKAP79 $_{19\text{-}51}$, $_{25\text{-}61}$ and $_{19\text{-}61}$, Figure 5.8.

Residue Number	Sequence	Number Amino Acids
31-52	----------KASMLCFKRRKKAAKALKPKAG---------	22
25-61	------AERQKEKASMLCFKRRKKAAKALKPKAGSEAADVARK	37
19-52	AEGSPGAERQKEKASMLCFKRRKKAAKALKPKAG---------	34
19-61	AEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVARK	43
	**********************	

Figure 5.8. Sequence alignment of the AKAP79 Calmodulin binding site. Showing three elongated sequences derived to increase the propensity of helix formation of the binding site. The red letters indicate the putatative 1-11 CaM binding motif.

Table 5.3. AKAP79 Calmodulin binding domain α-Helix prediction. Table shows the percentage propensity of each individual residue in a short peptide to form an α-helix as predicted by the AGADIR α-helix / coil transition theory algorithm. A value of greater than 2 % is a good indicator of helix formation (Munoz and Serrano, 1997).

Residue	Amino acid	% Helical content			
		19-61	19-52	25-61	31-52
19	A	0.0	0.0		
20	E	0.0	0.0		
21	G	0.0	0.0		
22	S	0.0	0.0		
23	P	3.2	3.2		
24	G	3.4	3.5		
25	A	6.4	6.4	0.0	
26	E	7.4	7.5	0.2	
27	R	8.6	8.6	0.7	
28	Q	8.9	8.9	1.3	
29	K	8.5	8.5	1.6	
30	E	8.1	8.1	1.8	
31	K	7.6	7.6	2.1	0.0
32	A	7.3	7.3	2.2	0.1
33	S	6.3	6.4	2.2	0.2
34	M	7.2	7.2	3.9	1.9
35	L	7.2	7.2	4.1	2.2
36	C	7.2	7.2	4.4	2.6
37	F	7.0	7.0	5.2	3.8
38	K	6.3	6.2	4.9	3.7
39	R	5.6	5.6	4.6	3.6
40	R	5.2	5.1	4.3	3.5
41	K	4.4	4.3	3.7	3.0
42	K	3.8	3.7	3.2	2.6
43	A	3.6	3.4	3.1	2.5
44	A	3.1	2.9	2.7	2.2
45	K	2.1	1.9	1.9	1.5
46	A	1.8	1.6	1.6	1.3
47	L	1.0	0.9	0.9	0.7
48	K	0.0	0.0	0.0	0.0
49	P	0.0	0.0	0.0	0.0
50	K	0.1	0.0	0.1	0.0
51	A	0.1	0.0	0.1	0.0
52	G	0.1	0.0	0.1	0.0
53	S	0.3		0.3	
54	E	1.1		1.1	
55	A	1.2		1.2	
56	A	1.3		1.3	
57	D	1.2		1.3	
58	V	1.2		1.2	
59	A	1.0		1.1	
60	R	0.7		0.7	
61	K	0.0		0.0	

5.2.2 AKAP79 CBD Cloning

Following the successful cloning and expression of the unstructured AKAP79 M a similar strategy was used for the AKAP79 CBD constructs. The three constructs, 19-51, 24-61, 19-61 were cloned into the pOPIN-S vector using the same method as AKAP79 M, outlined below.

Primers were designed that included a specific 15 bp linker sequence homologous to the vector needed for the ligation independent In-Fusion™ cloning (Clonetech). The AKAP79 CBD constructs were amplified with the primers using the AKAP79 FL as the template. The PCR mix was then incubated with Dpn I to remove any trace template before cleaning with a mini prep kit (QIAGEN). The AKAP79 CBD constructs DNA was then added with the pOPIN-S to the In-Fusion™ mix and heated at 42°C for 30 minutes before the reaction was stopped with cold TE buffer, and mixture transformed into XL-1 blue cells. The infusion cloning followed by blue-white screening gave a good mix of white and blue colonies on the agar plate; only white colonies were seen on the positive control. Subsequent PCR screening of the white colonies with the initial primers confirmed the presence of the construct. The colonies were then grown up in 10 ml LB overnight and the plasmid extracted by mini-prep (Qiagen). Samples of each plasmid were sent for Sanger sequencing (Geneservice) using stock T7 forward and reverse primers. The full sequencing confirmed the designed clones (see appendix 3).

5.2.3 AKAP79 CBD Expression and Purification

All three AKAP79 CBD were expressed and purified using the same procedures; outlined below is the procedure for $AKAP79_{19\text{-}61}$.

The $AKAP79_{19\text{-}61}$ Sumo construct was transformed into BL21(DE3) cells and expressed at 37°C, before induction at 18°C overnight with 1 mM IPTG. After cell lysis and centrifugation at 48000 g, the supernatant was applied to a Ni^{2+} Affinity column (5 ml HisTrap™ FF, GE), washed with 10 column volumes of His buffer (500 mM NaCl, 50 mM Tris, 5 mM DTT pH 8.0) and $AKAP79_{19\text{-}61}$ eluted with 250 mM imidazole. After buffer exchange using a HiPrep™ 26/10 (GE) desalting column to remove the imidazole, Sumo protease (0.2 mg) was added and cleavage allowed to proceed at 4°C for 18 hours. Following the cleavage the protein preparation was reapplied to the Ni^{2+} Affinity column. The flow-through which was cleaved AKAP79 $CBD_{19\text{-}61}$ was collected, concentrated and desalted using an Amicon® Ultra (Millipore, 3K MWCO). The concentrated $AKAP79_{19\text{-}61}$ was then further purified using HPLC on a C18 analytical column (4.5mm x 10 cm 5 µm resin diameter 100 Å pore width, Jones chromatography) with a 5 – 50% acetonitrile + 0.01% trifluroacetic acid (TFA) gradient. The final peptide peak was gyrovaped to remove acetonitrile before freeze drying to a powder to allow accurate mass measurement, which confirmed the peptide to be $AKAP79_{19\text{-}61}$.

To prepare uniformly ^{13}C,^{15}N labelled $AKAP79_{19\text{-}61}$ media containing ^{13}C- labelled glucose and ^{15}N ammonium chloride in a 2xM9 minimal media (MM) was used, using standard methods for, expression. The purification protocol was as described above.

5.2.4 AKAP79 CBD Results

5.2.4.1 Expression and Purification

The purification steps followed the same rational as with previously purified constructs - French press, centrifugation, HisTrap™ affinity column, desalt and overnight cleavage with Sumo protease. AKAP79$_{19\text{-}61}$ was cleaved overnight from Sumo at 4°C rather than at 30°C as was used for the preparation of AKAP79 M (Chapter 4). This gave near 100% cleavage, Figure 5.9. The difference in effectiveness at the lower temperature was thought to be caused by the use of a new batch of Sumo protease that had greater efficacy. A heat step was also included to denature any proteases that may have still been present. This was initially introduced after the initial centrifugation following cell lysis. This, however, led to a loss of approximately 20% of the yield. The heat step was then moved to after the Sumo cleavage with little loss in yield.

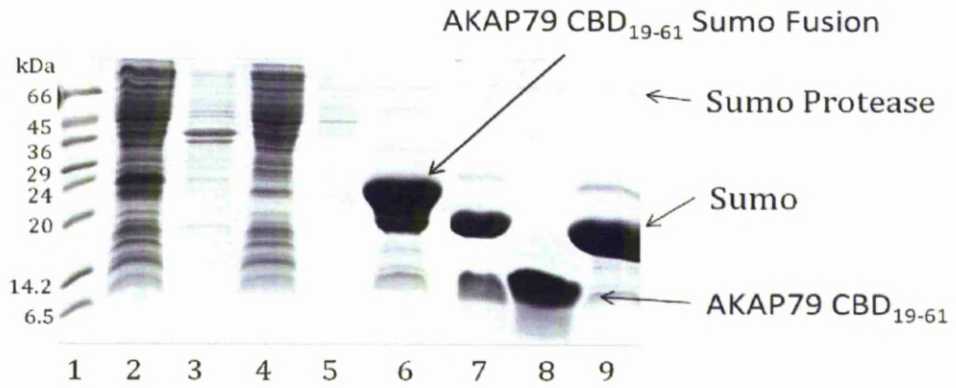


Figure 5.9. AKAP79 CBD$_{19\text{-}61}$ purification. 18 % SDS PAGE Gel shows: Lane1=markers, 2=post induction,3= pellet, 4=HisTrap™ flow through, 5=wash step, 6=the peak fraction after addition of 250 mM imidazole,7= post Sumo protease incubation at 4°C, 8=post second HisTrap™, post second 250 mM imidazole elution. The prominent band at approximately 14 kDa is AKAP79 CBD$_{19\text{-}61}$ post-cleavage.

Another problem faced was the possibility of disulphide bond formation between each peptide through a cysteine residue in the centre of the sequence, this led to an increase in DTT to 5 mM.

Initially to desalt and concentrate the peptide a combination of dialysis and freeze drying was used. Unfortunately, the dialysis membrane used (500 Da cellulose ester, Spectra/Por®, Spectrumlabs) did not reliably retain the peptide even though the MWCO was six times less than the peptides molecular weight. This was possibly due to the membrane which was supplied wet, drying out in places that allowed the escape of the peptide. The best technique for freeze-drying the peptide was to split the peptide solution into several microcentrifuge tubes before freeze-drying the samples. This also allowed for easy storage of the dried peptides. The peptides only provide a small change in absorbance at 280 nm and do not react to the Bradford assay, most likely due to them containing only one aromatic residue. This made it difficult to use a final ion exchange cleaning step because of the lack of signal. However, the nature of the unstructured peptide allowed the use of HPLC for purification where the absorbance was measured at 214 nm, due to the amide bond. A C5 analytical column was used which gave a good peak but on further inspection the shoulder was thought to contain other species of very similar hydrophobicity. A C18 analytical column was then used with a gradient 5 – 50% acetonitrile gradient. This chromatogram showed several separated peaks, Figure 5.10. The largest peak was identified as the $AKAP79_{19\text{-}62}$ by mass spectrometry.

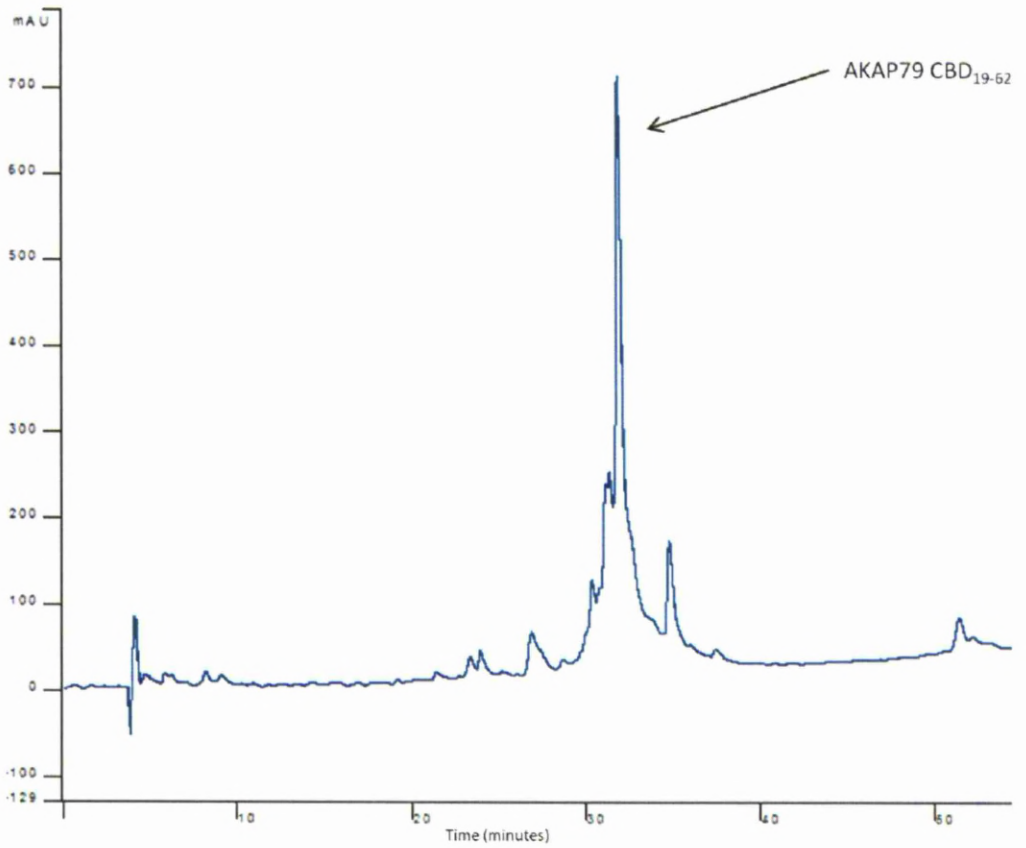


Figure 5.10. AKAP79 CBD_{19-62} purified by HPLC. Double labelled $^{13}C/^{15}N$ AKAP79 CBD_{19-62}.resuspended in dH_2O and 50µl loaded onto C18 analytical column with a 5 – 50% acetonitrile + 0.01% trifluroacetic acid (TFA) gradient. Main peak fraction was gyrovaped to remove acetonitrile and freeze dried.

All three AKAP79 CBD constructs expressed and purified well, but due to time constraints it was decided that further work would be carried out on $AKAP79_{19-62}$.

5.2.4.2 Circular Dichroism

CD was used to analyse the secondary structure of the peptide. When the peptide was in Tris buffer the (0% TFE) the peptide showed a partial α-helix depicted by the single minimal region at 208 nm followed by a rise. As the TFE was introduced

the sample converted to a predominately α-helical form as depicted by the two minimal regions at 208 and 225 nm. This occurred at just 20% TFE with little further changes at 40 and 50% TFE, Figure 5.11. The % helix formation for the peptide with the addition of TFE increased from -0.37 % (0 TFE) to ~ 2 % helix with between 20 and 50 % TFE.

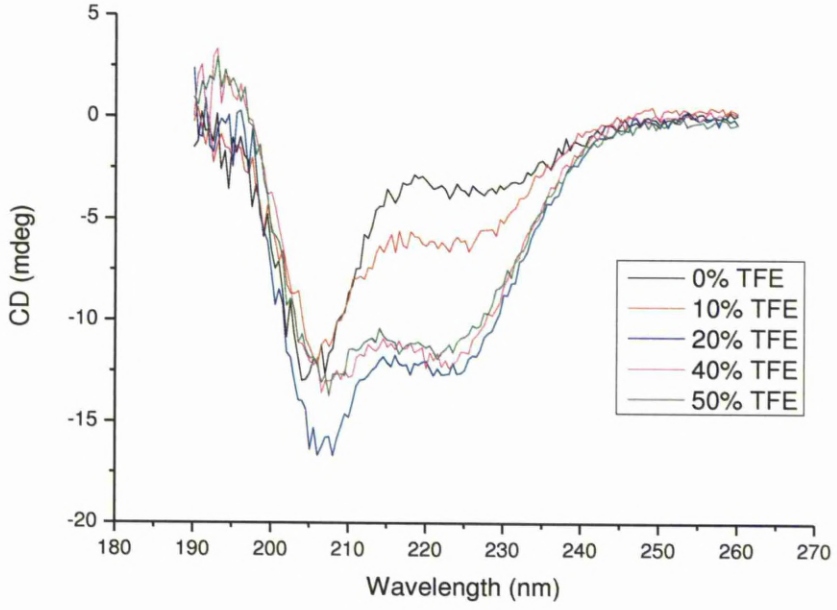


Fig 5.11. CD spectra of AKAP79$_{19-62}$. Recombinant peptide, 1 mM in 50 mM Tris buffer, pH 7.0 was titrated with 10 / 20 / 40 and 50% trifluroethanol, TFE. The Tris buffer was measured and subtracted. The % helix of the peptide was initially -0.37 % then increased to 0.49 % at 10% TFE before increasing to 2.5, 2.4 and 2.1 % for 20, 40 and 50% TFE respectively.

5.2.4.3 NMR Studies

Lyophilised $^{13}C/^{15}N$ $AKAP79_{19-62}$ was resuspended in 50 mM Tris 5 mM DTT pH 6.5 buffer and 2D ^{15}N HSQCs were taken. Unfortunately, the spectra were poor with variation in peak sizes and degradation seen. The 2D ^{15}N HSQCs were repeated using a new preparation of peptide but more poor data was generated with fewer peaks also registered. It was thought that there may have been a problem with amino acid clipping of the peptide or possible problems with the HPLC fractionation. Further analysis showed that the cysteine present (C_{36}) was being gradually oxidised over the course of the NMR experiments (as assessed from the ^{13}C chemical shifts of the cysteine C^{β}) and hence degrading the NMR spectrum. This also explained the array of peaks in the HPLC purification of the peptide.

This was later proved correct by re-cloning the peptide with the cysteine mutated into a serine ($C_{36}S$) (this work carried out by other members of the laboratory while this thesis was being written).

5.3 Discussion

CaM has been shown to be a dynamic protein that can drastically change its conformation from a collapsed form to an extended helix with the addition calcium (Ishida et al., 2002). Upon binding with peptides large conformational changes in CaM have also been seen such as with CaMKII where both lobes collapse around the ligand (de Diego et al., 2010; Vetter and Leclerc, 2003). From the CaM ^{15}N HSQC spectra presented here the chemical shift changes indicate that the conformational change where the lobes collapse around the peptide is most likely. From the chemical shift mapping presented, the AKAP79 calmodulin binding domain appears to be able to interact with both the N-terminal and C-terminal domains of CaM. The model derived from de Deigo, kuper et al. 2010 supports this, showing the mapped chemical shift changes in the C-terminal lobe and the linker helix that are positioned in proximity to the modelled peptide helix.

The $AKAP79_{31-52}$ does not conform to the more common CaM binding motifs but to a novel 1-11 binding motif recently found in calcineurin calmodulin binding domain, CnA-CBD (Majava and Kursula, 2009). Interestingly the CnA-CBD was shown to interact with CaM in a 1:1 and 2:2 ratio. Present experiments cannot rule out the formation of a 2:2 ratio for $AKAP79_{31-52}$, although a 1:1 ratio is more likely further experiments are required.

To map exactly the interaction site of AKAP79 to CaM would require further NMR experiments with double labelled AKAP79 peptide. This was not achieved in the time available due to unforeseen problems with the cysteine residue of the

recombinant peptides. The cloning, expression and purification showed that the methods worked well and the production of peptides was possible.

One caveat however, is that even with both components – CaM and AKAP79 peptide – labelled, there will still be problems in using NMR for the high-resolution structural work since severe line-broadening in the spectrum of the complex might preclude sufficient data from being acquired for the full structural calculations.

The CD of the recombinant $AKAP79_{19-62}$ peptide was unaffected by the cysteine as the experiments were carried out at much lower concentrations than the NMR experiments and over a shorter timescale. The CD data showed that the secondary structure of the peptides was an α-helix as proposed.

Future work would continue with the NMR study of the CaM and AKAP79 interaction using recombinant mutant peptides. A crystallisation study might also be possible and/or binding interactions analysed via SAXS to show any conformational changes.

CHAPTER 6

BIOPHYSICAL CHARACTERISATION OF SAP97

6.1 Introduction

SAP97 is a large multi domain protein that acts as a scaffold protein at the cell membrane. It contains, from the N-terminus, the L27 domain, three PDZ domains and a fused a src homologue 3 (SH3)-guanylate kinase like (GK) domain which is separated by a flexible Hook region, Figure 6.1.1. Previous structural characterisation of SAP97 included solved structures of the three PDZ domains (Cabral et al., 1996; von Ossowski et al., 2006; Wang et al., 2005; Zhang et al., 2007) and the L27 domain (Feng et al., 2004). The structure of SH3-GK of SAP97 has proved elusive, possibly due to difficulties in crystallising the domains caused by a longer Hook region than that of the closely related PSD-95 that has two solved structures (McGee et al., 2001; Tavares et al., 2001).

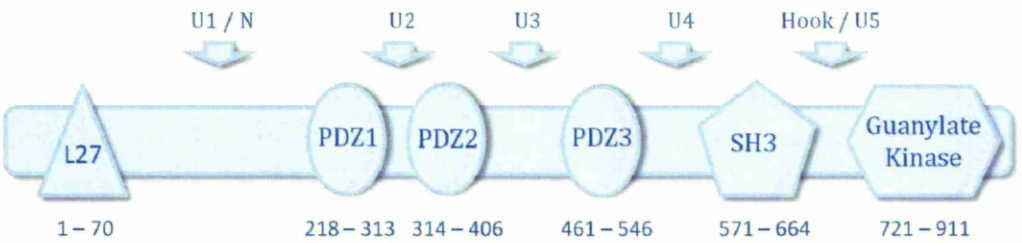


Figure 6.1.1. Schematic of SAP97 showing the locations of each domain. Individual domains are joined together by linker regions termed U regions that are unstructured.

It is unlikely that SAP97 is just a collection of discrete domains. Rather, interdomain interactions exist which are important for the function of SAP97. Previous work by our group investigated the conformation of the triple PDZ region (Goult et al., 2007). A computational model reported by Wu et al 2000, which used the high resolution structures of similar domains from other molecules predicted fairly extensive conformational changes upon interactions with GKAP

(Wu et al., 2000). Low resolution EM images showed SAP97 to exist as a mixture of bent or elongated forms (Nakagawa et al., 2004).

The aims here are to characterise SAP97 further, using several constructs which express multiple and single domains. Initial characterisation were done using Size-Exclusion Chromatography-Multi Angle Light Scattering (SEC-MALS) and NMR. Chapter 7 describes in more detail the structural characterisation using SAXS. A summary of the multi domain constructs made are in Figure 6.1.2.

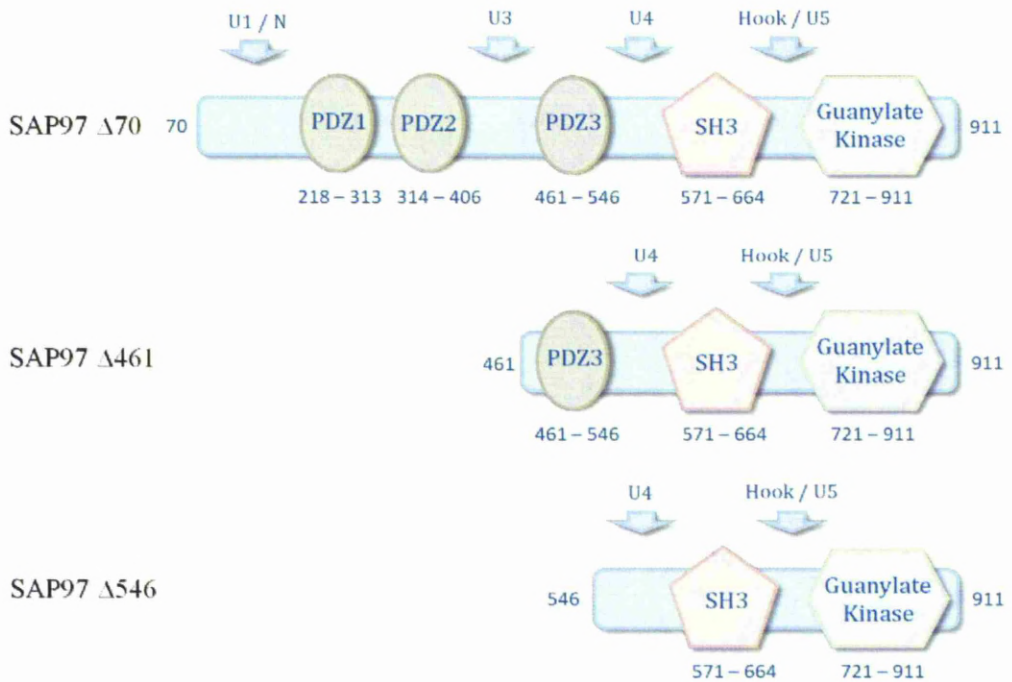


Figure 6.1.2. Schematic of SAP97 Constructss. Each construct was created to characterise SAP97 overall structure. Each construct was cloned into pLEICS-01 vector.

A full length SAP97 was not created because of concerns about the oligomerisation effects the L27 domain might have on the protein. Previously the L27 domain of SAP97 was shown to form a heterotetrameric complex with the L27 domain of Lin-2 and was still able to interact and bind to further L27 domains

to form large clusters of complexed proteins (Feng et al., 2004). Due to the potential complications this might create, it was decided to create a SAP97 construct without the L27 domain.

6.2 SAP97 Δ70, Δ461 and Δ546

6.2.1 Cloning

All the cloning of the SAP97 constructs was carried out by our collaborator Dr Mark Leyland, University of Leicester, into the vector pLEIC-01. This ampicillin resistant vector contains a hexahistag followed by a TEV protease cleavage site which allows removal of the tag after protein purification.

Table 6.1 Summary of all the SAP97 constructs used in this thesis.

Protein	Construct	Vector
SAP97 Δ70 Mr = 94kDa	HexaHis-TEV-70 – 911	pLEIC-01
SAP97 Δ461 Mr = 53 kDa	HexaHis-TEV-461 – 911	pLEIC-01
SAP97 Δ546 Mr = 46 kDa	HexaHis-TEV-546 – 911	pLEIC-01
SAP97 GK Mr = 23.5 kDa	HexaHis-TEV-721 – 900	pLEIC-01
SAP97 N PDZ1 Mr = 29 kDa	HexaHis-TEV-70 – 313	pLEIC-01

6.2.2 Expression and purification

The three constructs were expressed and purified using the same conditions, unless stated.

For unlabelled, rich media preparations the plasmids were transformed into BL21(DE3) cells (Novagen) and grown in 1 litre LB containing 0.1 g/L Amp until 0.7 – 0.9 OD_{600} (30°C, 200 rpm). The cells were induced with 1 mM IPTG and left to grow overnight at 18°C, 200 rpm. Cells were pelleted at 11000 g for 15 minutes, resuspended in His buffer (0.5 M NaCl, 50 mM Tris-HCl, 2 mM DTT, pH 8) and 1 EDTA free protease inhibitor tablet (Roche) and 250 ng DNase I and stored at -20°C overnight.

The cells were thawed, lysed by French press at 1000 PSI and centrifuged at 48000 g for 30 minutes. The resultant supernatant was collected and filtered through 0.22 µm syringe filter ready for chromatography

A three-step purification scheme was used for all three constructs. The clarified supernatant was loaded onto a 5ml HisTrap™ High Performance (GE Healthcare). The column was washed with increasing concentrations of imidazole (10, 20, 50 mM), at 2 ml/min flow rate, 6 column volume (CV) at each increment. The protein was eluted with 250 mM imidazole and buffer exchanged using HiPrep™ 26/10 desalting column (GE Healthcare) to remove imidazole ready for the next purification step.

A 5 ml HiTrap™ Q FF anion exchange column (GE Healthcare) was next. The sample was loaded, washed and then eluted using a 0 – 500 mM KCl gradient over 100 ml at 2 ml/min flow rate. Following concentration to 5ml using an Amicon® Stirred Cell (10k MWCO) to approximately 5 ml the SAP97 samples were loaded onto a 26/60 Superdex™ 75 gel filtration column (Superdex™ 200 used for SAP97 Δ70) pre-equilibrated with Gel Filtration (GF) buffer (500 mM NaCl, 50 mM Tris-HCl, 2 mM DTT, 1 mM EDTA pH 7.5). Isocratic elution using 1 CV GF buffer at 1mm/min flow rate yield the desired protein. The peak fractions were concentrated with Amicon® filter (10k MWCO). The protein samples were deemed to be >95% pure using SDS page gel. Protein samples were identified by mass spectrometry.

To prepare uniformly ^{2}H, ^{15}N labelled SAP97, media containing ^{15}N ammonium chloride in a 2xM9 minimal media (MM) prepared in D_2O rather than H_2O was used (See appendix 1 for details of media). A single colony was grown in 5 ml LB for 4 hours; the sample was spun down, the LB media removed and the cells resuspended in 20 ml of unlabelled MM and grown to an OD_{600}, of 0.6. The media was removed and the cells resuspended in 100 ml ^{15}N MM in D_2O. The cell culture were allowed to grow for some time, diluted two fold, allowed to grow again followed by further additions of D_2O until a total of 1 litre of D_2O was added, the cells were allowed to grow until an OD_{600} of 0.5-0.6, 1 mM IPTG added and expression allowed to proceed overnight at 18°C. The purification was performed identically to the unlabelled preparations as described above. The SAP97 protein samples were concentrated using an Amicon® Ultra (Millipore,

3K MWCO) to 1 mM. All NMR experiments were at acquired 298K on the Bruker 800 MHz spectrometer.

6.2.3 Results

The plasmids designed and made by Dr Mark Leyland were transformed in BL21(DE3) (Invitrogen) cells and the respective proteins expressed. An approximate yield 20 mg/l of protein was purified from the rich LB media. The final purity is as shown in Figure 6.2.1. One problem that arose through purification was that the hexahistag could not be removed by Tev protease. Several concentrations and temperatures were used but the constructs would degrade before complete cleavage. It was decided to use the constructs with the hexahistag intact than to risk degradation and decrease the yields to an unusable level.

Degradation of the expressed proteins was a problem. Steps such as increasing the reducing agent such as DTT or introducing extra purification steps such as ion exchange on top of gel filtration did not stop the degradation. This could be seen in the samples after the NMR ^{15}N-^{1}H HSQC experiments (800 MHz, 298K, 50mM Tris 500 mM NaCl, pH 7.0), Figure 6.2.2, where extra bands on the gel indicated the presence of proteins at lower molecular weights. The degradation might be caused by the constructs containing unstructured flexible N-terminal domains or the elongated Hook domain found in SAP97. Ultimately, it might be necessary to re-engineer the constructs to totally resolve the problem. Fortunately,

the degradation was not too severe to prevent further analysis of the expressed proteins.

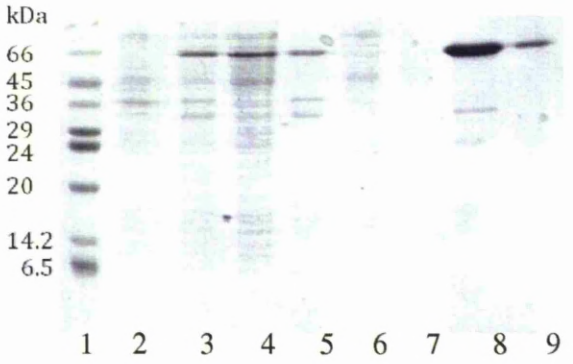


Figure 6.2.1. 10% SDS PAGE gel showing purified SAP97 Δ461. The lanes are Low range markers (Sigma), SAP97 Δ461 pre induction, post induction, lysate, pellet, HisTrap™ flow through, wash, 250 mM imidazole elution and elution post gel filtration respectively. White mark shown was caused during scanning of the gel after destaining: no bands were masked.

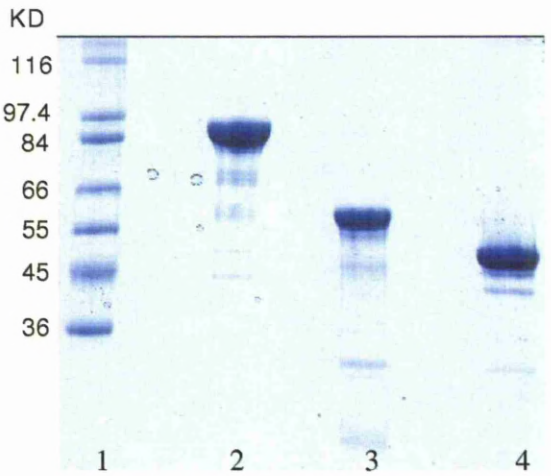


Figure 6.2.2. 10% SDS PAGE gel showing purified SAP97 proteins. The lanes are: (1) High range markers (Sigma), (2) SAP97 Δ70, (3) Δ 461 and (4) Δ 546. Each sample was taken after the NMR experiments and showed a very small degree of degradation.

6.2.4. NMR Characterisation

A purified sample of each of the three SAP97 constructs was buffer exchanged into 50 mM Tris, 500 mM NaCl, pH 7.0 and concentrated to 0.2 mM before insertion into an NMR tube and data collected on the spectrometer. The initial spectra of ^{15}N-labelled samples gave very poor signal to noise. This is most likely caused by the large size of the proteins. The constructs were then re-expressed in deuterated medium. This has the effect of improving the resolution and the sensitivity of the spectra. For proteins of this size, the recommend experimental technique is based on the TROSY experiment. The experiment overcomes the line broadening effects due to rapid transverse relaxation, thereby giving spectra which have narrow line widths. Figure 6.2.3 to 6.2.5 show the ^{15}N-^{1}H TROSY spectra of the ^{15}N/^{2}H labelled samples. Although the growth media was D_2O, the amide protons are back-exchanged during the purification stages which were all performed using H_2O buffers.

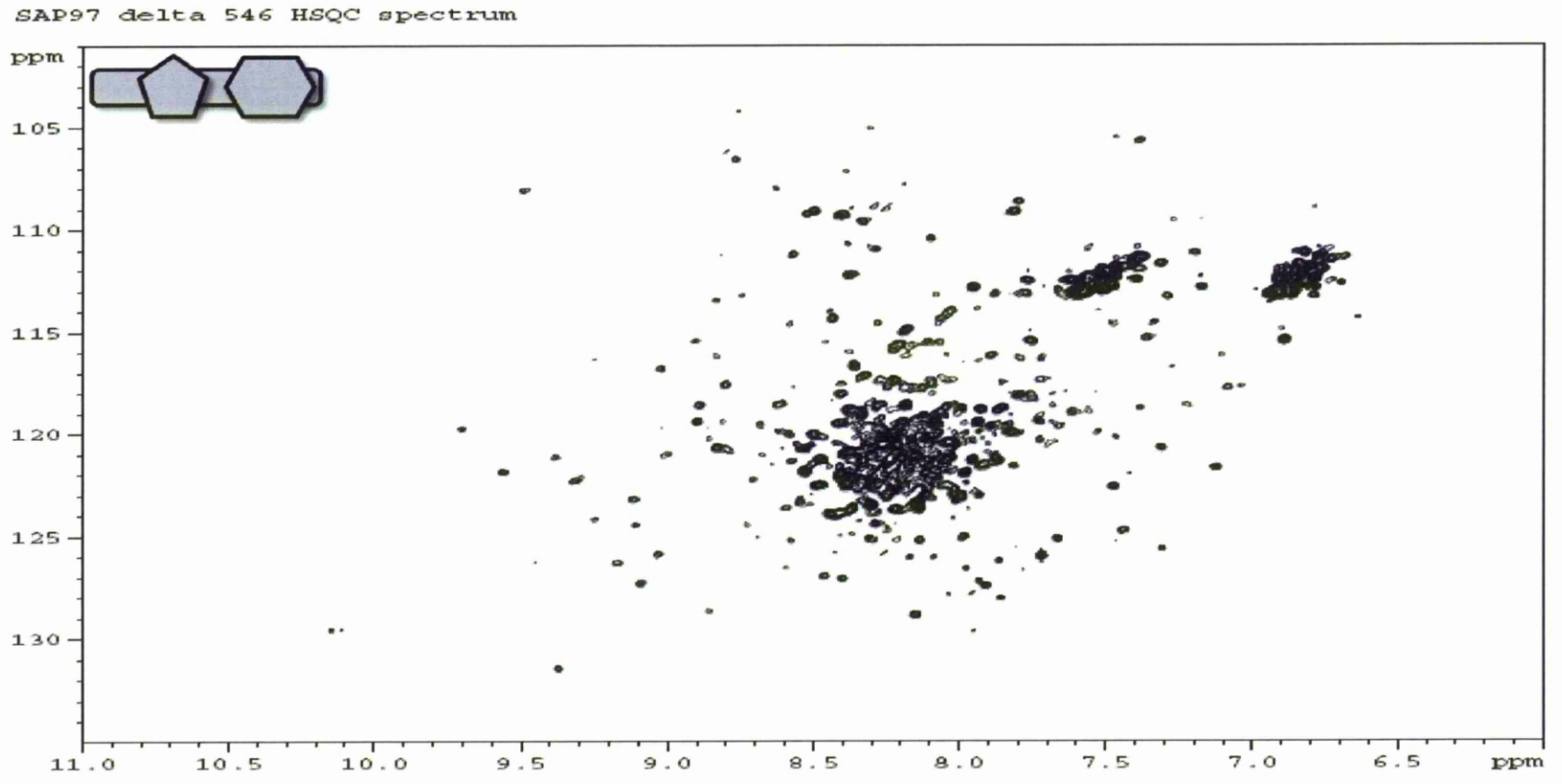


Figure 6.2.3. **^{15}N-^{1}H HSQC spectrum of SAP97 Δ546 at 298K 50mM Tris 500 mM NaCl, pH 7.0.**

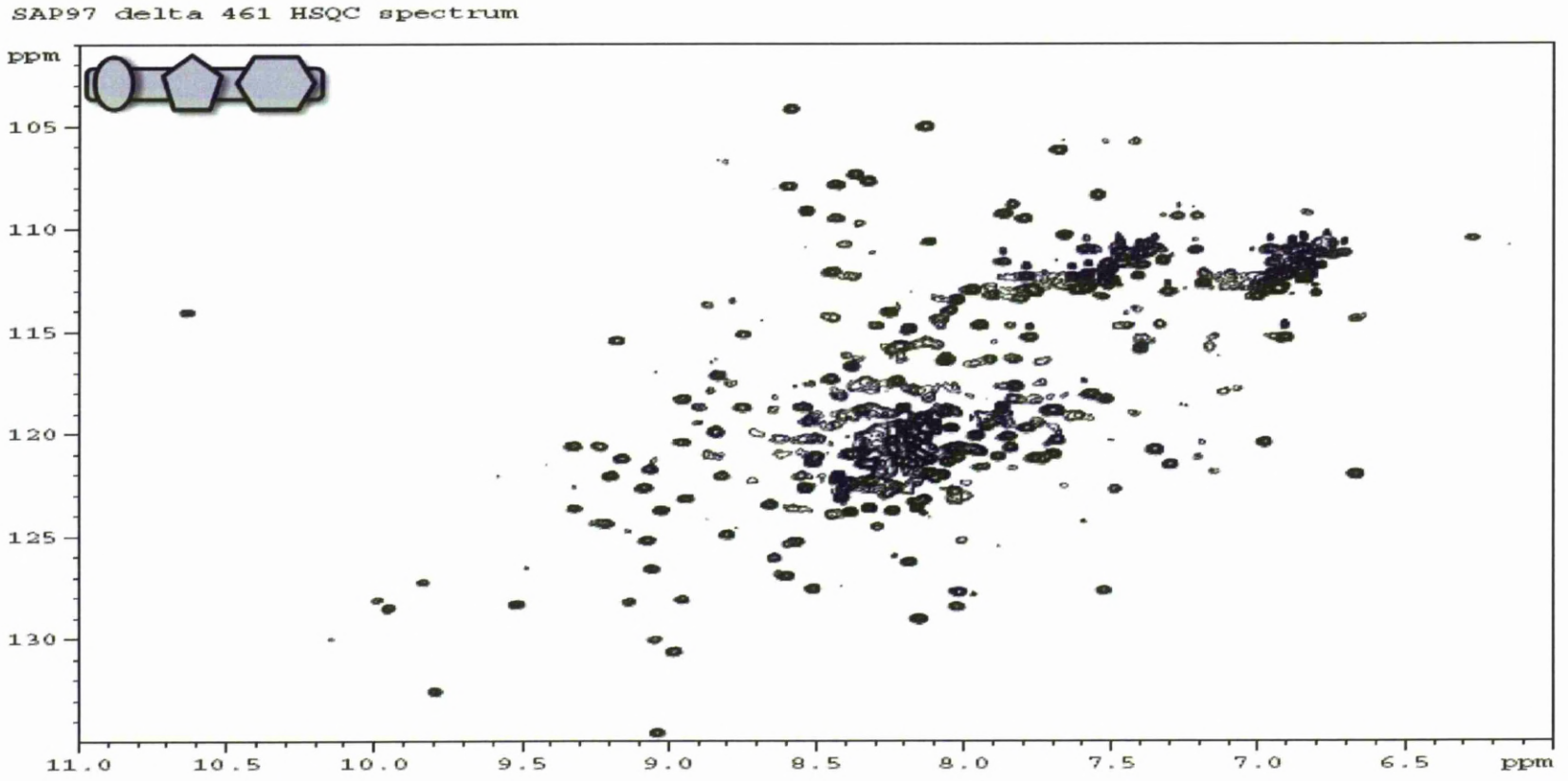


Figure 6.2.4. ^{15}N-^{2}H HSQC spectrum of SAP97 Δ461 at 298K 50mM Tris 50 mM NaCl, pH 7.5.

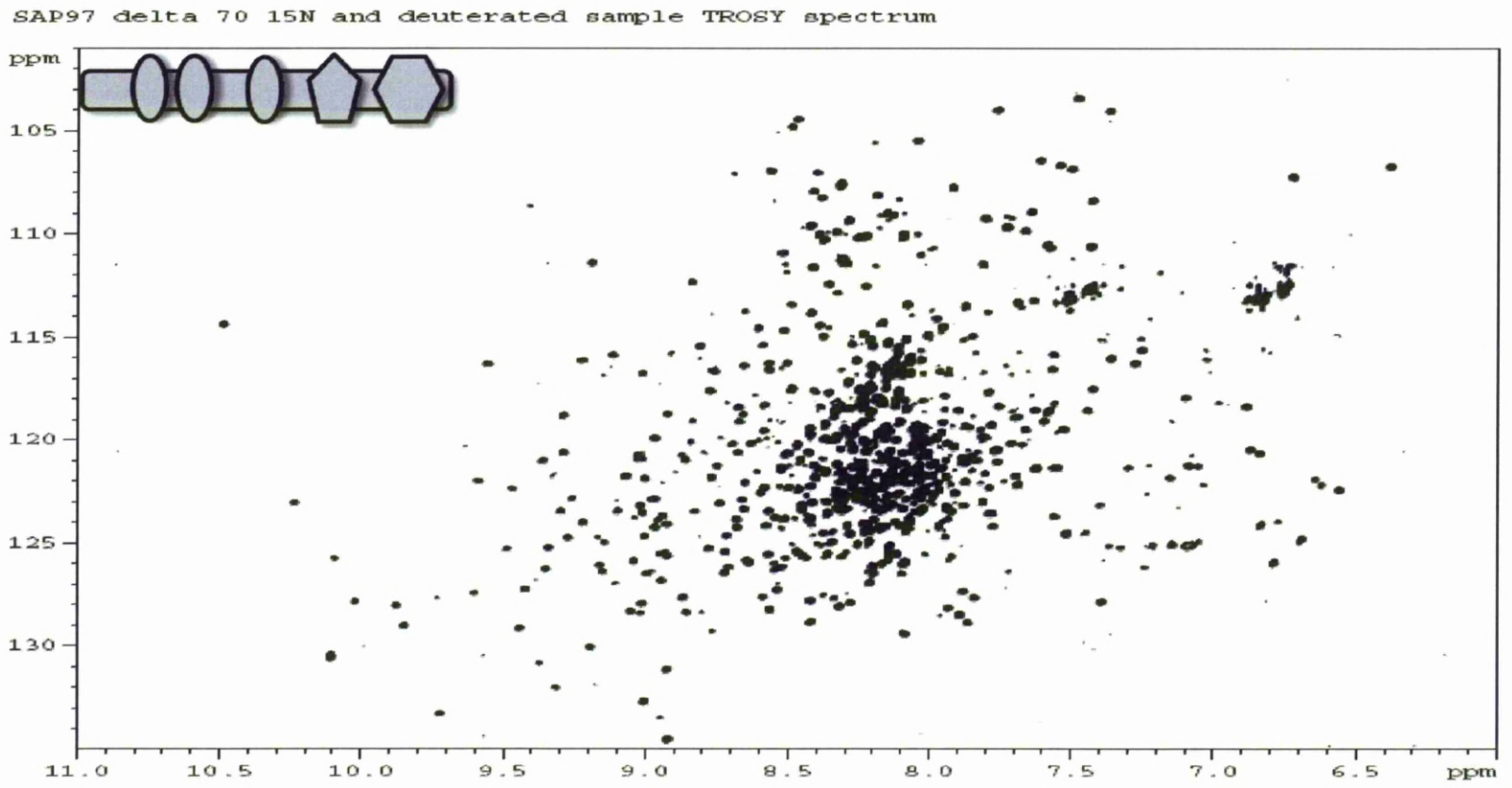


Figure 6.2.5. ^{15}N-^{2}H TROSY spectrum of SAP97 Δ70 in at 298K 50mM Tris 500 mM NaCl, pH 6.5

To achieve deuteration, the culture media and protein production protocol was modified. As described in the methods section, the initial stages took approximately 15 hours to reach the induction stage but did give a very good level of deuteration (approximately 85%) although the yield was lower than when using the protonated media, at approximately 10 mg/l of culture.

The ^{15}N-^{1}H HSQC spectrum of SAP97 Δ546 showed a large number of peaks bunched together in the centre of the spectra that indicates an unfolded or disordered area. This is most likely caused by the U4 linker at the N-terminal end of the Δ546 and the U5 / Hook domain that separates the SH3 and GK domains.

The ^{15}N-^{1}H TROSY spectrum of deuterated ^{15}N-labelled Δ461 is shown in Figure 6.2.6. The residues from the different domains are easily identified by overlaying with the separate spectra of PDZ3 and of Δ546. The ^{15}N-^{1}H spectrum of PDZ3 was previously assigned (Goult et al., 2007). Assigning the PDZ3 section within Δ461 is a major undertaking which is beyond the scope of this thesis. In this work, the PDZ3 resonances from Δ461 were assigned using the 'nearest neighbour' approach, that is, the peak from the spectrum of Δ461 which is closest to the peak from the isolated PDZ3 was tentatively assigned to a PDZ3 residue within Δ461. This method is an accepted method when comparing spectra in the absence of complete sequential resonance assignment. The current comparison suggests that PDZ3 is mobile and has limited contacts with the SH3-GK domain of SAP97. This approach

of comparing resonance from different domains and ascribing which residues arise from which domain will be used throughout this thesis.

The ^{15}N-^{2}H TROSY spectrum of Δ70 shows only approximately half of the expected 805 peaks (847 total residues – 43 prolines residues which have no amide protons). The good chemical shift dispersion of many of the resonances indicate the presence of folded domains (Figure 6.2.5) When the spectra was overlaid with both the PDZ1 N domain and SAP97 PDZ123 (Goult et al., 2007) approximately 95 % of the peaks matched (Figure 6.2.7 left). This interestingly suggests that the residues missing from the Δ70 spectra belong to the SH3-GK domain. In addition, the good match of the resonances from the PDZ123 domain in the spectra of PDZ123 and Δ70 suggests that the SH3-GK domains do not interact with the PDZ123 domains to cause changes in conformation.

When Δ70 was overlaid with the individual PDZ domains (spectra previously acquired by other members of the group) the spectra did not align completely. Correcting for differences in pH, this results suggests that within the PDZ region, intramolecular interactions exist.

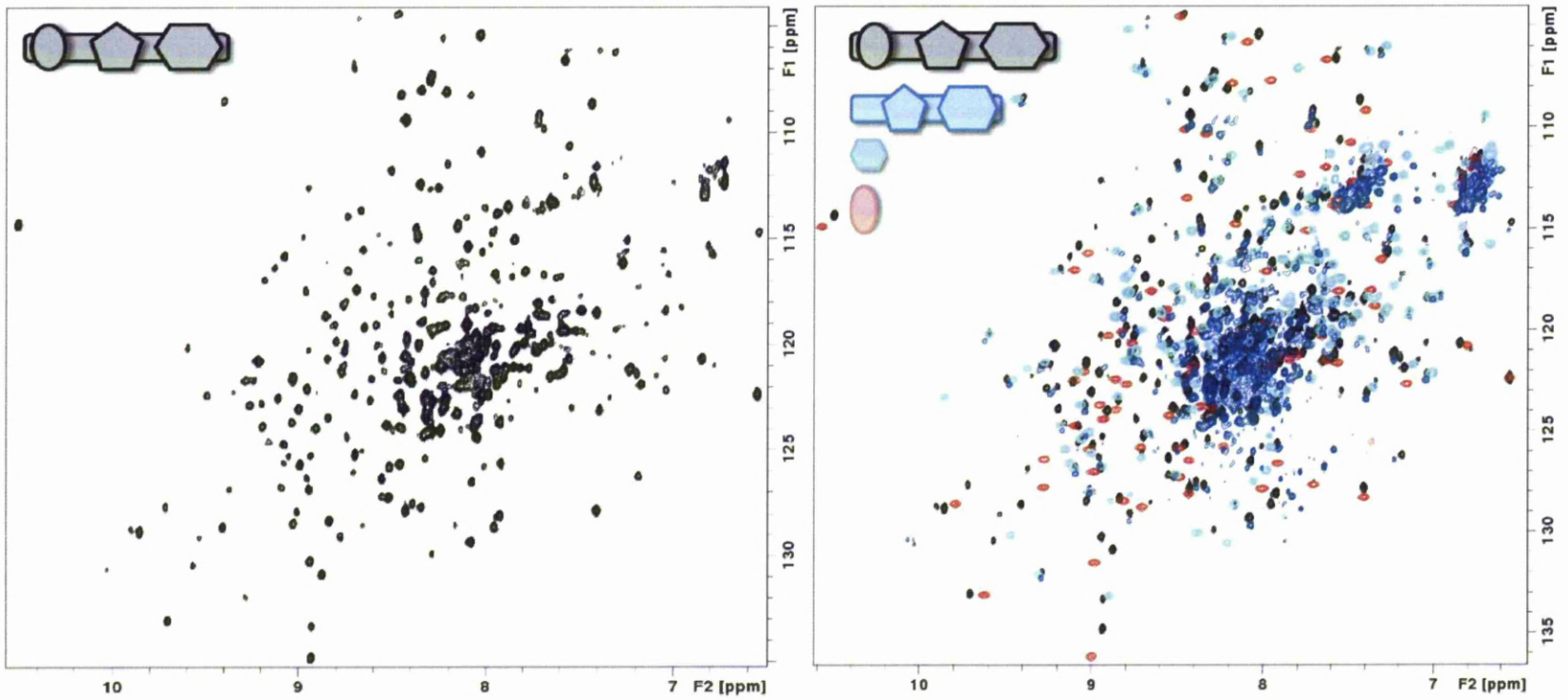


Figure 6.2.6. ^{15}N-^{1}H TROSY spectrum of ^{2}H/^{15}N-SAP97 Δ461 at 298K 50mM Tris 50 mM NaCl pH 7.5 alone (Left) and overlay of SAP97 Δ461 (black) with SAP97 Δ546 (blue), SAP97 GK (cyan) and PDZ3 individual domain (red).(Right).

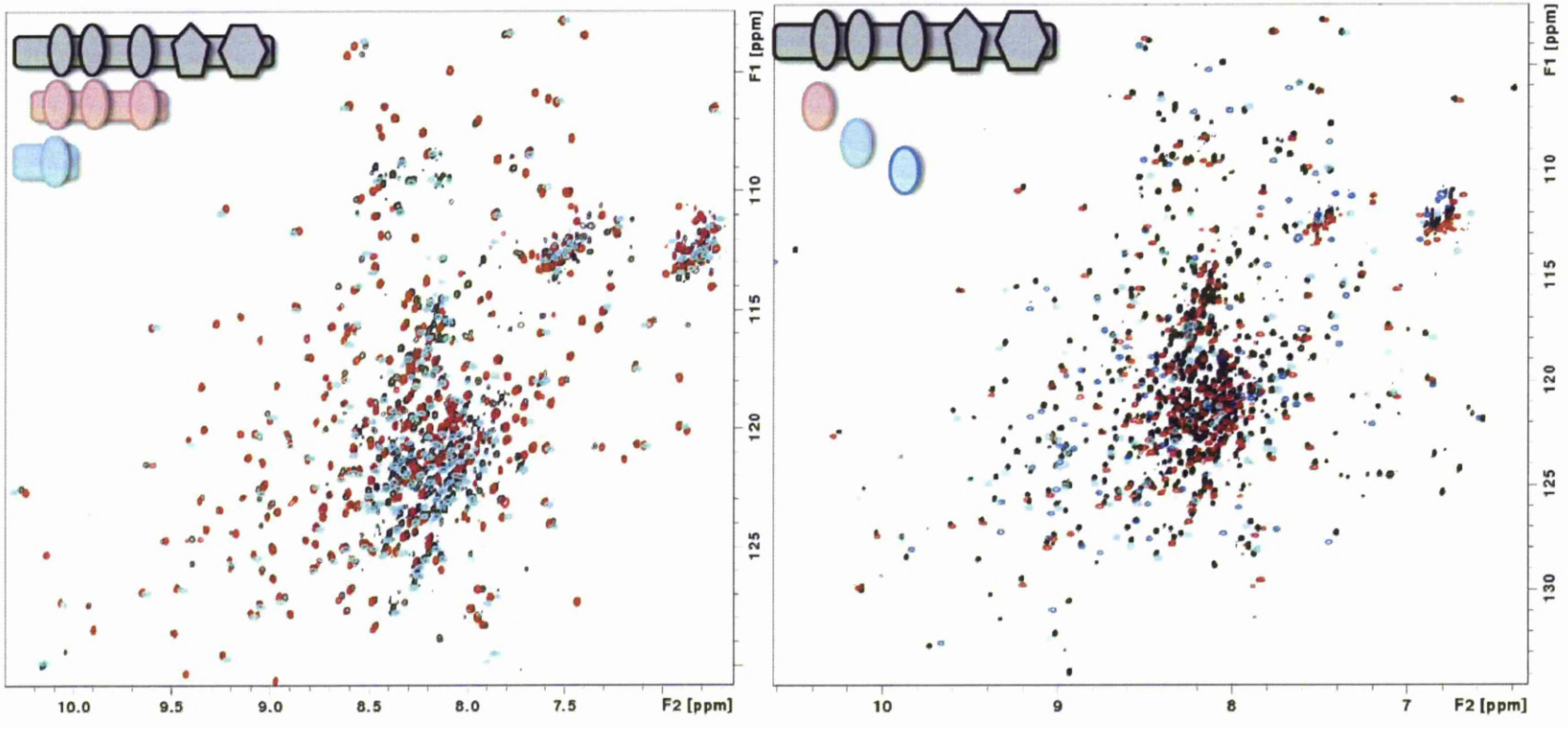


Figure 6.2.7. ^{15}N-^{2}H TROSY spectrum of SAP97 Δ70 in at 298K 50mM Tris 500 mM NaCl, pH 6.5. Left, overlaid with PDZ123 (red) and SAP97 N PDZ1 (cyan), showing that most of the visible peaks belong to the PDZ domains. Right, ^{15}N-^{2}H TROSY spectrum of SAP97 Δ70 overlaid with individual PDZ 1,2,3 spectra (red, cyan and blue respectively); the peaks do not match as well as when using the intact triple PDZ domain.

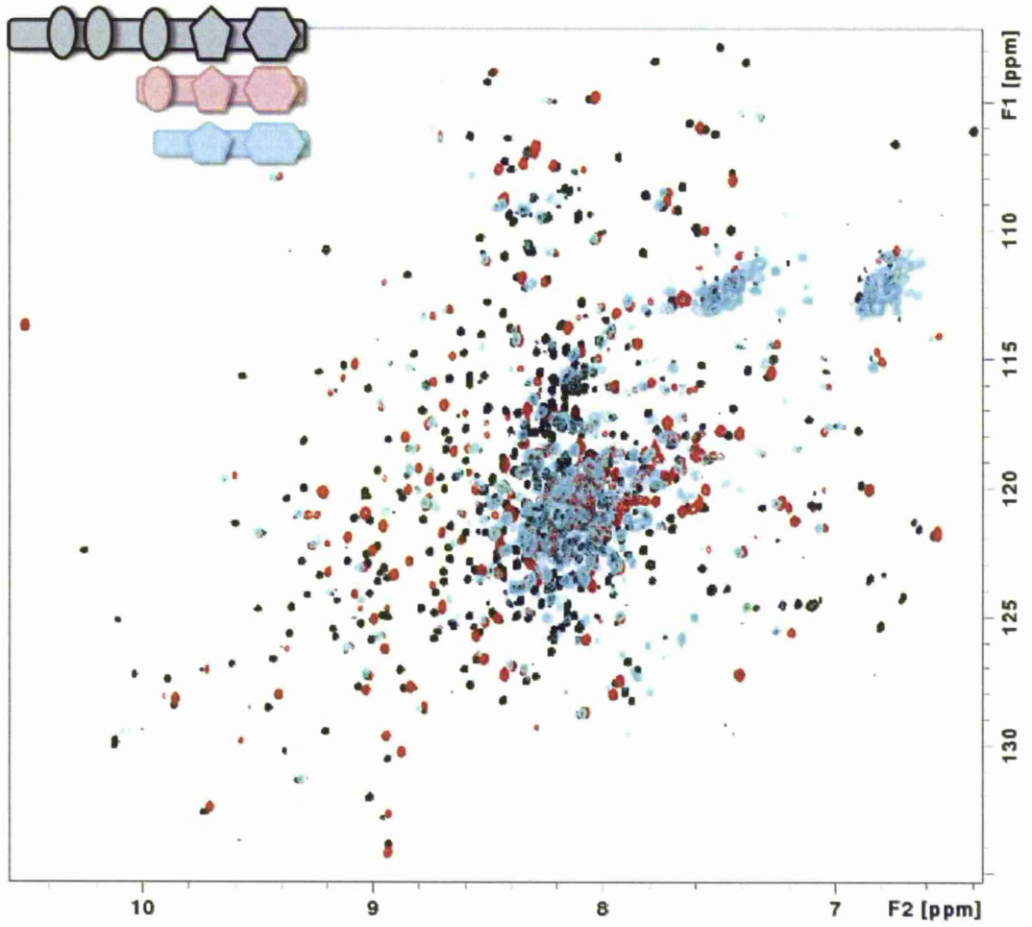
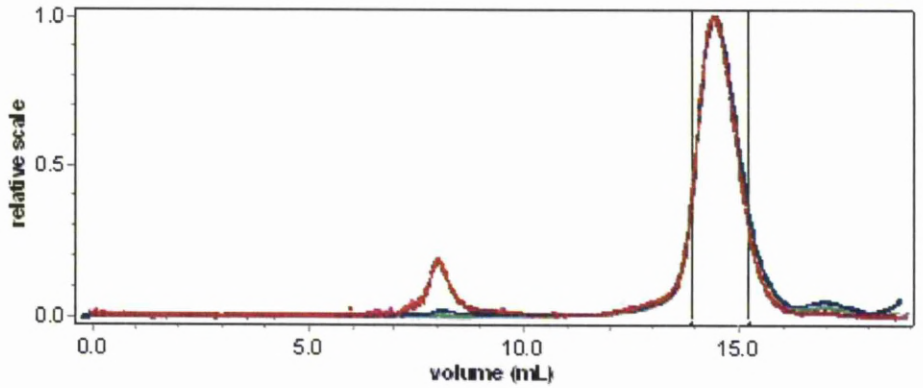


Figure 6.2.8. ^{15}N-^{2}H TROSY spectrum of SAP97 Δ70 in at 298K 50mM Tris 500 mM NaCl, pH 6.5. Δ70 (black) overlaid with Δ461 (red) and Δ546 (cyan). Little movement between Δ70 and Δ461 with most prominent peaks, 546 overlaps well with 461, and unstructured area in centre is possibly Hook region.

When SAP97 Δ70 was overlaid with Δ461 and Δ546 (Figure 6.2.8) the prominent peaks from Δ546 matched up well to the Δ461 peaks which in turn matched up well to the Δ70 peaks. Taken together with the above comparison with the PDZ123 spectrum, the data suggests that the PDZ domains have little, and probably transient interactions, with the SH3 – GK domain.

6.2.5 SAP97 SEC-MALLS Characterisation

Each of the purified SAP97 constructs was subjected to size exclusion chromatography – multi angle laser light scattering, (SEC-MALLS, Dawn DSP, Wyatt Technology). The scattering data was processed using Dynamics^{TM} and this gave the full molecular weight of each construct as 93.7 kDa, 53.3 kDa and 46.7 kDa and a hydrodynamic radius of 4.3, 2.1 and 2.0 nm for Δ70, Δ461 and Δ546 respectively. This suggests that Δ70 is twice as long as the other constructs and that Δ461 and Δ546 are very similar in size. The SEC-MALLS also demonstrated that constructs are predominantly monomeric, Figure 6.2.8.



	Δ70	Δ461	Δ546
Molar mass moments (g/mol)			
Mn	9.412e+4	5.392e+4	4.683e+4
Mp	9.504e+4	5.542e+4	4.741e+4
Mw	9.425e+4	5.409e+4	4.690e+4
Mz	9.438e+4	5.426e+4	4.696e+4
M(avg)	9.372e+4	5.323e+4	4.667e+4
rms radius moments (nm)			
Rn	n/a	n/a	n/a
Rw	n/a	n/a	n/a
Rz	n/a	n/a	n/a
R(avg)	n/a	n/a	n/a
Hydrodynamic radius moments (nm)			
Rh(n)	4.3	6.5	8.2
Rh(w)	4.3	6.6	8.0
Rh(z)	4.3	6.8	7.8
Rh(avg)	**4.3**	**2.1**	**2.0**
Translational diffusion moments (cm^2/sec)			
Dt(n)	5.69e-7	1.07e-6	1.10e-6
Dt(w)	5.69e-7	1.07e-6	1.10e-6
Dt(z)	5.68e-7	1.07e-6	1.11e-6
Dt(avg)	5.61e-7	1.59e-8	2.68e-8

Figure 6.2.8. Top SEC-MALLS elution profile for SAP97 Δ461. Table of SEC-MALLS data for SAP97 Δ70, Δ461, Δ546.

6.2.6 SAP97 Δ70, Δ 461 and Δ 546 – Interactions with Calmodulin

Calmodulin, CaM, is a Ca^{2+} binding protein that acts a secondary messenger within the cytoplasm that has previously been shown to bind to SAP97 by surface plasmon resonance, SPR (Paarmann et al., 2008). To characterise our constructs further the interactions between each construct and CaM were investigated by fluorescence spectrophotometry. Both CaM and the SAP97 have no intrinsic fluorescence (Vorherr et al., 1990). To impart excitation fluorescence, CaM was bound to dansyl chloride which fluoresces at 340 nm. The presence to dansyl chloride does not affect the binding properties of native CaM (Malencik et al., 1982).

The fluorescence intensities of dansylated CaM often varies from preparation to preparation. This was due to the presence of free dansyl chloride. To resolve this, extensive dialysis was used with an initial two hour, room temperature dialysis followed by a second overnight dialysis at 4°C, both in 20 mM ammonium bicarbonate, pH 7.5.) Emission Spectra for Δ70 show an increase in absorbance with increased Δ70 and a small shift to lower wavelengths (blue shift). The blue shift is most likely caused by intrinsic tryptophan emission from Δ70, Figure 6.2.9.

The fluorescence data yielded three binding curves (Figure 6.2.10) and from these the K_d value for each binding reaction was determined (Table 6.2. The binding data for Δ546 gave a very large standard error. This was due to increasing dansyl CaM fluorescence with each repeat assay. A possible reason for this was free dansyl still present in the sample. From the binding curve for Δ461 it is noticeable that a greater

concentration of Δ461 was required to reach saturation. In each case complete saturation was not achieved, this could lead to discrepancies with the B_{max} and the K_d values.

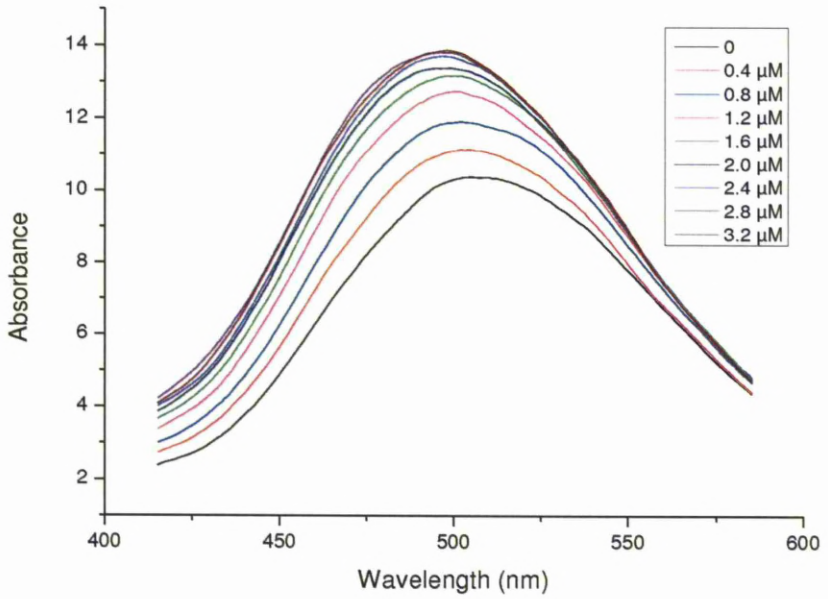


Figure 6.2.9. Emission Spectrum of dansyl calmodulin after complex formation with SAP97 Δ70. Excitation was performed at 340 through a slit width of 5 nm with a path length 1 cm. Titration of increasing concentration SAP97 Δ70 (µM) into 5 µM D-CaM. Complex formed in the presence of Ca^{2+}.

Table 6.2. The kinetic characteristics of the interaction with dansyl CaM with SAP97, Average of three repeats using a one-site saturation curve, $Y = (B_{max} \times X) / (K_d + X)$. The equilibrium dissociation constant K_d is given.

SAP97 Construct	K_d (µM) n=2
Δ546	3.21 ± 0.8
Δ461	27.2 ± 1.5
Δ70	1.01 ± 0.2

The equilibrium dissociation constant K_d from the SPR was 122 ± 46 nM for a SAP97 Δ575 construct (Paarmann et al., 2008) compared to the 3.21 ± 0.8 μM for Δ546. It is not clear why there is such a large difference between the two studies; it is possible that the tethering of the SAP97 Δ575 construct in the SPR altered the conformation of the construct opening the binding site to allow the CaM to bind more freely. Or the dansylation of the CaM in the fluorescence experiments inhibited the binding to SAP97; saturation in the titrations was also not achieved that could also lead to an inaccurate calculation of K_d; more investigations are required to resolve these differences. The addition of a PDZ3 domain in Δ461 showed a further decrease in binding affinity 27.2 ± 1.5 while Δ70 showed the greatest binding affinity 1.01 ± 0.2 μM. Using dansylated CaM for these experiments have proved not to be too reliable; further studies of the interactions between CaM and SAP97 proteins were abandoned at this stage until a more reliable method could be found.

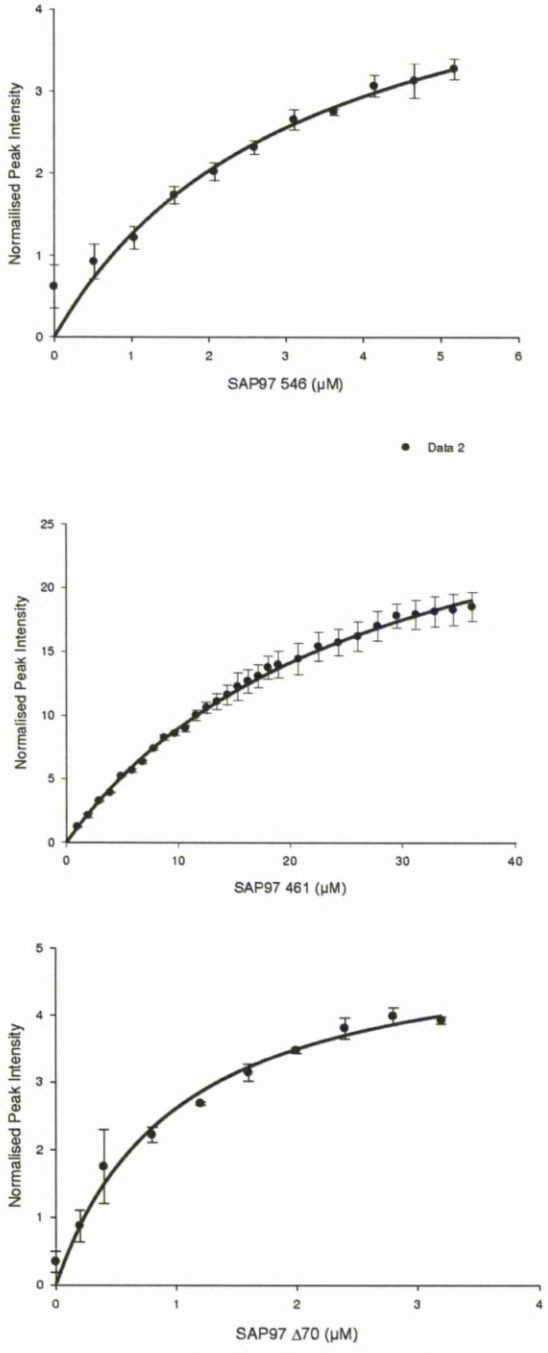


Figure 6.2.10. Fluorescence measurements of SAP97 and Calmodulin interaction. Top graph shows the normalised peak intensity of dansyl calmodulin binding to a varying SAP97 Δ546 concentration, error bars show standard error. Middle graph shows the normalised peak intensity of d-CaM binding to a varying SAP97 Δ461 concentration, error bars show standard error. Bottom graph shows the normalised peak intensity of d-CaM binding to a varying SAP97 Δ70 concentration, error bars show standard error.

6.3 SAP97 GK Domain

A new construct of SAP97 GK domain was produced by Dr. Mark Leyland from the University of Leicester. The expectation is that the GK domain would give improved NMR data which would be suitable for further structural studies.

6.3.1 SAP97 GK Expression and Purification

The expression and purification of SAP97 GK domain was similar to the other SAP97 constructs.

For unlabelled, rich media preparations the plasmid was transformed into BL21(DE3) cells (Novagen) and expressed in 1 L LB plus 0.1 g/l Amp until 0.7 – 0.9 OD_{600} (30°C, 200 rpm). The cells were induced with 1 mM IPTG overnight at 18°C, 200 rpm. Cells were pelleted at 11000 g for 15 minutes and resuspended in His buffer (0.5 M NaCl, 50 mM Tris-HCl, 2 mM DTT, pH 8) and 1 EDTA free protease inhibitor tablet (Roche) and 250 ng DNase I and stored overnight at -20°C.

The cells were thawed and lysed by French press at 1000 PSI and centrifuged at 48000 g for 30 minutes. The resultant supernatant was collected and filtered through 0.22 μm syringe filter ready for chromatography

A two-step protocol was used for the purification of SAP97 GK. The supernatant was loaded onto a 5ml HisTrap™ Fast Flow (GE Healthcare) and washed with increasing concentrations of imidazole (10, 20, 30 mM), at 2 ml/min flow rate, 6 CV at each increment. Proteins were eluted with 250 mM imidazole and the recovered sample from the elution buffer exchanged using HiPrep™ 26/10 desalting column (GE Healthcare) to remove imidazole. SAP97 GK was further purified using a 5 ml HiTrap™ Q FF anion exchange column (GE Healthcare). The sample was loaded, washed and then eluted with a 0 – 500 mM KCl gradient over 100 ml at 2 ml/min flow rate. The protein was then concentrated using Amicon® Ultra (Millipore, 10k MWCO).

To prepare uniformly ^{13}C,^{15}N labelled SAP97 GK, media containing ^{13}C- labelled glucose and ^{15}N ammonium chloride in a 2xM9 minimal media (MM) was used (See appendix 1 for details of media). The overnight step started with 1 colony grown in 1 ml LB for 4 hours before removal of the LB and resuspension in 40 ml MM overnight at 37 °C. The cells were centrifuged at 5000 g for 10 minutes to remove the media before whole cell pellet used to seed 1 L MM. It was necessary to increase the cell density before seeding the MM or no growth would occur but this had to be balanced against exposure of the cells to LB; the greater the exposure the greater the decrease in isotopic labelling of the protein.

The expression and purification was performed identically to the unlabelled preparations in accordance to the methods above.

6.3.2 SAP97 GK Results

SAP97 GK was expressed satisfactorily and purified to five >95% pure protein (Figure 6.3.1). A $^{13}C^{15}N$ labelled protein was expressed and purified. However, problems were encountered during the buffer exchange stage. At pH below 7.0 the protein has a tendency to crash out of solution, forming white precipitates. At pH above 7.0, the protein does not appear to be fully folded. The expressed GK domain also had a solubility issue in that although soluble at low concentrations, it had a tendency to precipitate at concentrations above 1mM. As the pI of the GK domain is 6.05 the poor solubility at pH greater that 7 is a surprise. The most likely explanation for this is that the absence of the SH3 domain destabilises the GK domain on its own. At the C-terminal of the GK domain there is a β-sheet that forms part of the SH3 fold. It is possible that this split domain also confers stability to the GK domain. Furthermore, the crystal structure of the SH3-GK domain from PSD95 (McGee et al., 2001) showed that there are important hydrophobic interactions between the SH3 and the GK domain. Absence of the SH3 domain removes these stabilising hydrophobic interactions as well as exposes a hydrophobic surface on the GK domain. Both these factors can explain the observation made here of poor protein solubility at high concentrations.

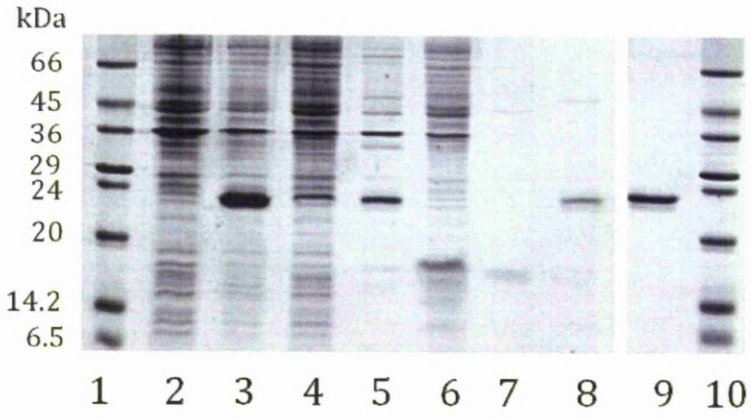


Figure 6.3.1. 15% SDS PAGE gel showing purified SAP97 GK. The lanes are (1) Low range markers (Sigma), (2)SAP97 GK pre induction, (3) post induction, (4) pellet, (5) lysate, (6) HisTrap™ flow through, (7) 10 mM imidazole, (8) 250 mM imidazole elution, (9) elution post ion exchange HiTrap™ Q FF and (10) Low range markers respectively. Lanes 9 and 10 were from a second SDS PAGE gel.

Although the protein was fragile a $^{13}C^{15}N$ sample was made. The ^{1}H HSQC spectrum was acquired at 600 MHz (Figure 6.3.2). This spectrum showed a good distribution of peaks which reflects that the protein was folded. What was also clear was that the protein was degrading over time in the spectrometer as seen by the resonances around 8ppm/128ppm (Figure 6.3.2). This degradation occurred even when the temperature of the NMR experiment was lowered to 295K. The degradation was such that 3D NMR experiments were not feasible and thus no structure determination was possible.

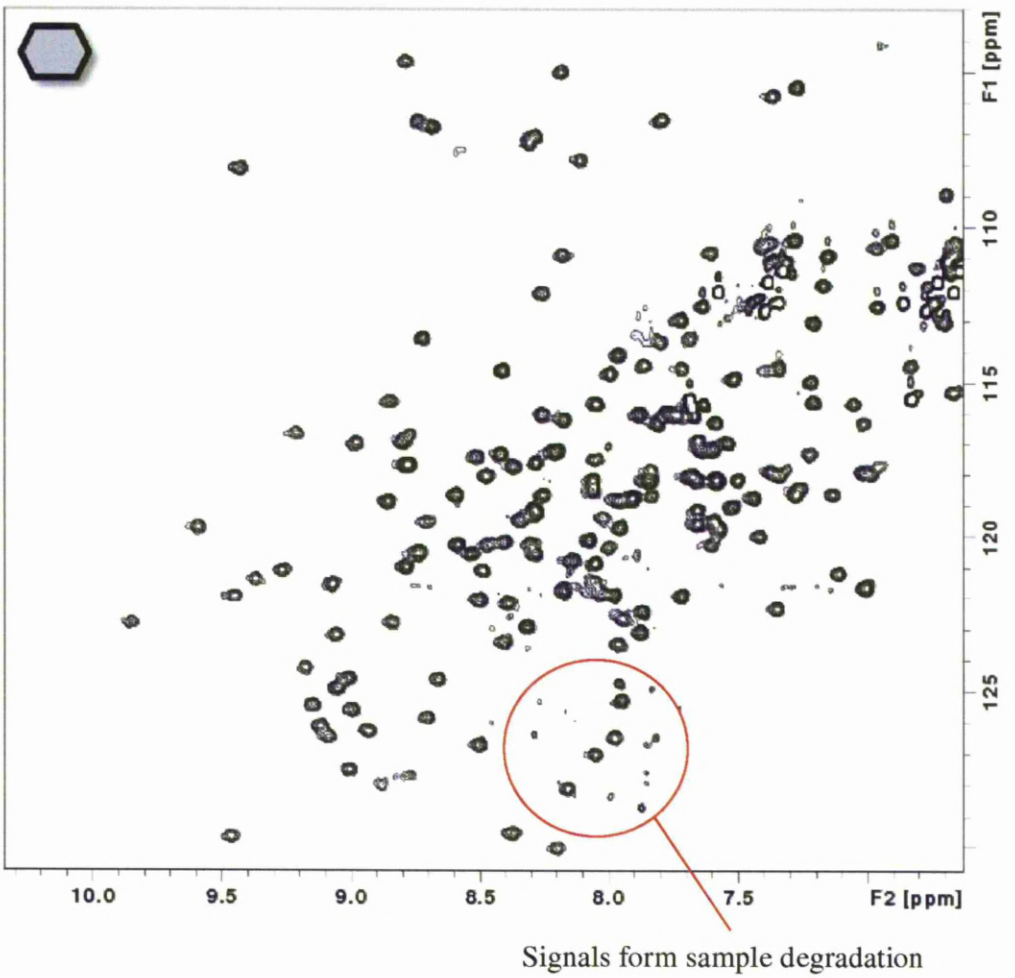


Figure 6.3.2. ^{15}N-^{1}H HSQC spectrum SAP97 GK domain. 600 MHz 50 mM NaCl 50 mM Tris HCl pH 7.5, 295 K. Red circle indicates resonances from degraded protein fragments.

6.3.3 SAP97 GK : GMP Binding

The GK domain in yeast has catalytic activity, catalyzing the reversible phosphoryl transfer from ATP to GMP.

The GK domain had been shown to undergo conformational change when in complex with GMP (Blaszczyk et al., 2001). Although it is recognised that the GK domain in MAGUK's is non-catalytic, it was wondered whether GMP could still bind to the GK domain from SAP97. A titration of GMP (Sigma) into the GK domain was conducted in the 800 MHz spectrometer (Figure 6.3.3) and specific chemical shifts were observed, demonstrating that the GK domain was able to interact with the GMP.

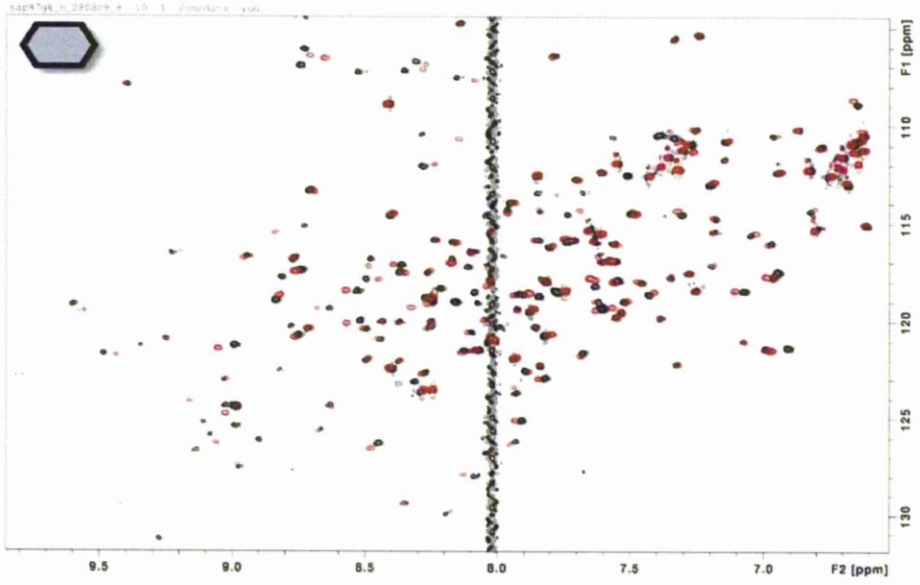


Figure 6.3.3. ^{15}N-^{1}H HSQC spectrum SAP97 GK : GMP complex. 800 MHz 50 mM NaCl 50 mM Tris HCl pH 7.5, 295 K. SAP97 GK in red and SAP97 GK : GMP complex (0.8 mM GMP) in black. The broad line at ~8.0 ppm is the resonance from the GMP.

6.3.4 Comparison of SAP97 GK domain with SAP97 Δ546 and Δ461

The GK domain is not a stable protein, due possibly to the absence of the split domain that it shares with the SH3 domain. To see if this had any large effect on the structure of the GK domain the ^{1}H HSQC of the $^{13}C/^{15}N$ labelled GK domain was overlaid over both the SAP97 Δ546 and Δ461 domains which both contain the SH3 and GK domains (Figure 6.3.4). Only about a handful of the resonances overlaid well in the spectra of the GK, Δ546 and Δ461. For the remaining resonances, there were considerable differences in shifts. What is clear from the overlaid spectra is that the GK domain does not contribute to the unstructured areas seen in Δ546 and Δ461. The significant differences in shift between the isolated GK domain and the GK domain in Δ564 and Δ461 suggests that the presence of the SH3 domain affects the magnetic environment around the GK domain. Removal of the SH3 is also likely to cause some conformational changes of the GK domain. The NMR results are not surprising since the SH3 domain is intimately linked to the GK domain, with considerable degree of interdomain interactions. It must be added here that resonances from the SH3-GK domain in Δ461 are weak in comparison with the resonances from the PDZ domain and Hook domains due to a higher degree of mobility in the PDZ and Hook domain. The intensities of the resonances from these two mobile domains are such that they visually mask the signals from the SH3-GK region.

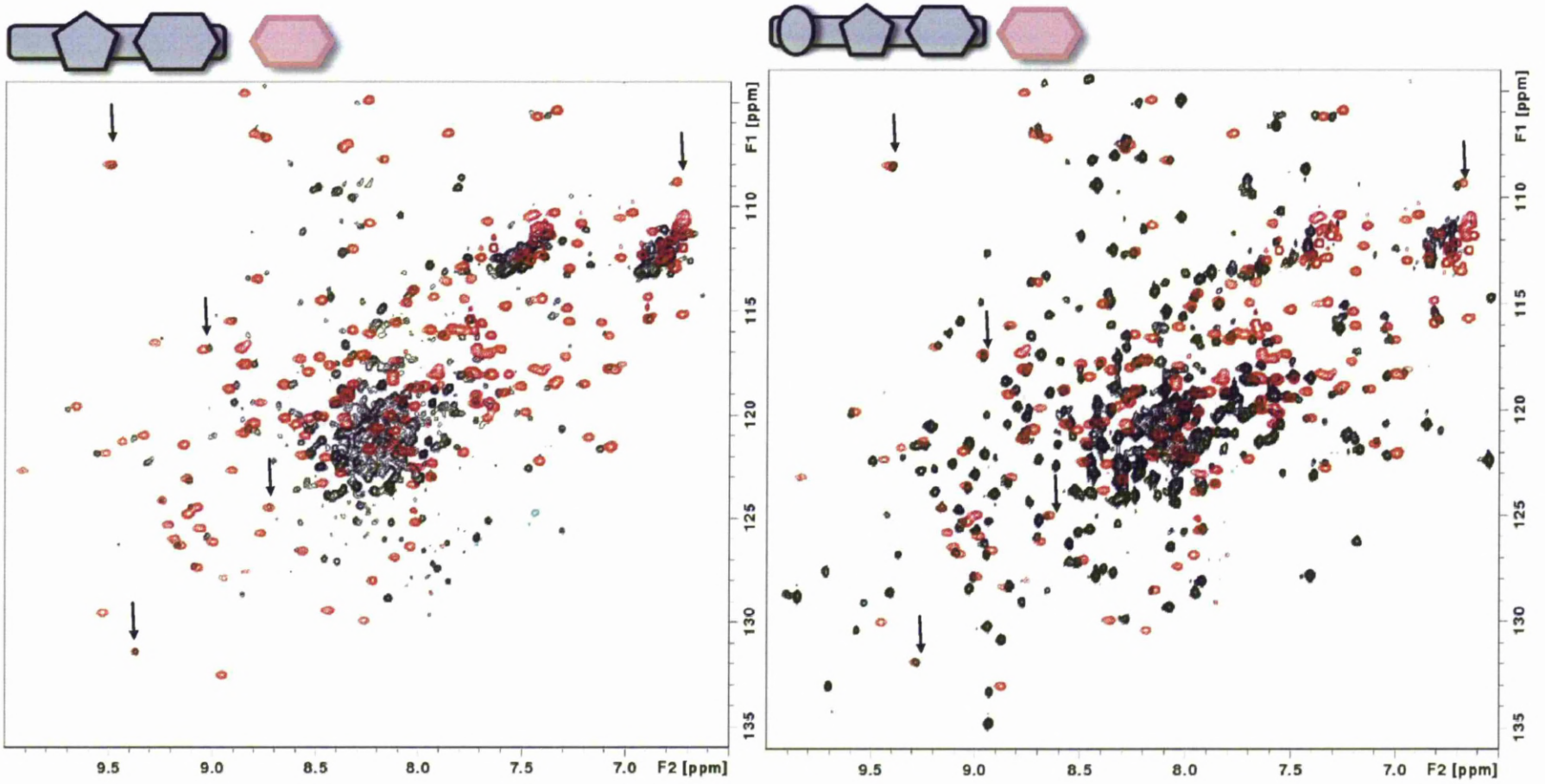
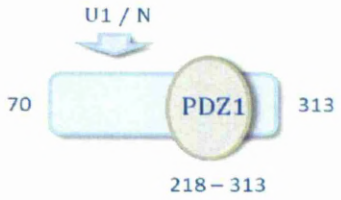


Figure 6.3.4. SAP97Δ546 and Δ461 overlaid with SAP97 GK. Left, ^{1}H HSQC spectrum of ^{13}C/^{15}N SAP97 GK domain (red) overlaid SAP97 Δ546 (black). Right ^{1}H HSQC spectrum ^{13}C/^{15}N SAP97 GK domain (red) overlaid SAP97 Δ461 (black). Arrows show peaks that match in both spectra.

6.4 SAP97 N-PDZ1 Domain

To investigate the interactions between the N domain (U1 linker region) and PDZ1 a new construct was created my Dr Mark Leyland from 70 – 318 that contains both PDZ1 and the N Domain.



6.4.1 SAP97 N-PDZ1 Expression and Purification

The expression and purification of SAP97 N-PDZ1 domain was identical to the SAP97 GK domain (see section 6.3.1.).

6.4.2 Results

The expression and purification were unproblematic and yielded approximately 15 mg/ml. A ^{15}N labelled purification was produced, Tev protease cleaved and buffer exchanged into 50 mM naCl, 50 mM Tris HCl, pH 6.5 and analysed by ^{15}N-^{1}H HSQC spectrum on the 600 MHz spectrometer.

The ^{15}N-^{1}H HSQC spectrum is shown in Figure 6.4.1. The plots are of the same spectrum plotted at different contour levels. The right hand figure clearly shows a

mixture of narrow and broader line widths with chemical shift characteristics to suggest that SAP97 N-PDZ1 has two domains - a folded and an unfolded domain. The unfolded region is identified by the bunching up of the resonance around 8-8.5ppm. The spectra of the PDZ1 domain overlay well with the dispersed resonances found in the spectrum of SAP97 N-PDZ1.

The data suggests that 150 or so amino acid downstream of the PDZ1 domain do not adopt a folded conformation; rather this region of SAP97 is unstructured. When the single SAP97 PDZ1 domain was overlaid with SAP97 NPDZ1 almost all of the peaks lined up suggesting that the N domain is not interacting with the PDZ domain.

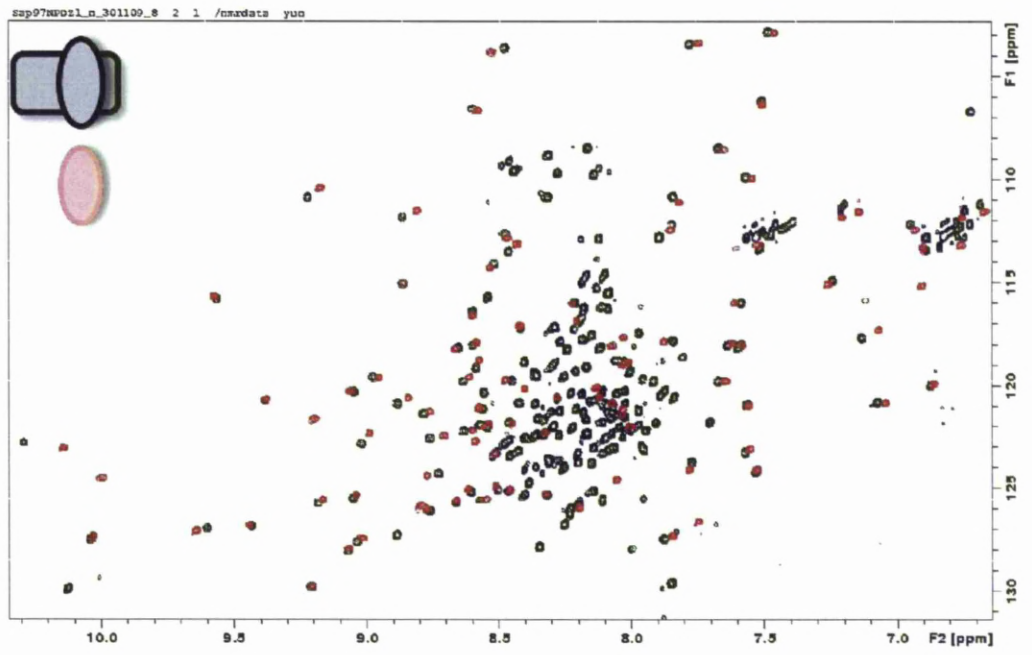


Figure 6.4.2. **^{15}N-^{1}H HSQC spectrum SAP97 NPDZ1 domain overlaid with PDZ1**. The SAP97 NPDZ1 (black) is overlaid with PDZ1 (red).

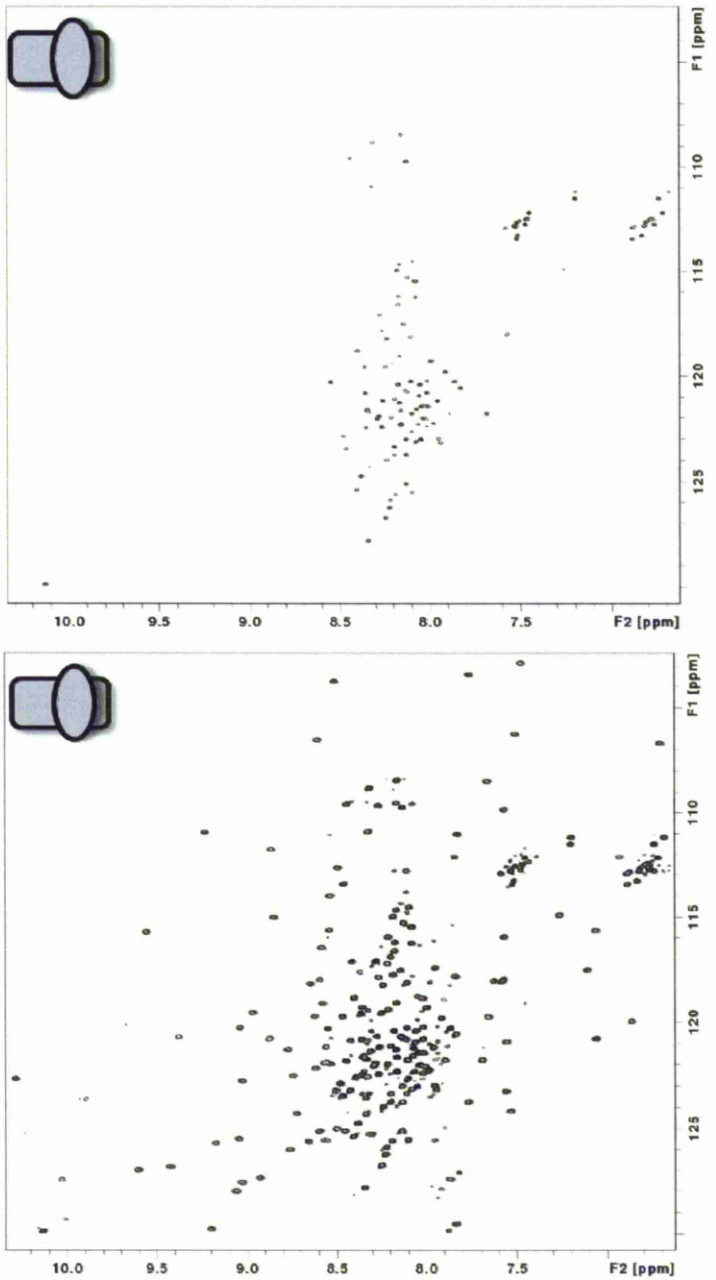


Figure 6.4.1. ^{15}N-^{1}H HSQC spectrum SAP97 NPDZ1 domain. The two plots are at different contour levels to emphasise the presence of two domains. The top plot, plotted at higher threshold levels has poor proton chemical shift dispersion and these signals must be from the unfolded domain upstream of the folded PDZ1 domains. The spectrum was acquired at 600 MHz, 50 mM Tris,50 mM NaCl pH 6.5.

6.5 SAP97 PDZ 2 Domain Interactions with Kir2.1 C-Terminus Domain.

Class I PDZ domains recognize the carboxyl terminal motif S/T- X - Φ, where X is any amino acid and Φ is a hydrophobic amino acid. In work carried out in our lab the three PDZ domains of SAP97 were each able to bind independently to the 10 a.a. C-terminal peptide, from Kir2.1 ^{418}EPRPLRRESEI^{428} . Both NMR and fluorescence spectrophotometry showed that PDZ2 binds with higher affinity than the other two PDZ domains (Goult et al., 2007). The peptide sequence contained the classical PDZ binding motif; the aim in this study is to investigate precisely how many amino acids in the C-terminus are in fact involved with PDZ interactions, and what additional residues might play a role in determining ligand specificity. To test this, a DNA construct expressing the last sixty amino acids of Kir2.1 was created by our collaborator, Dr. Mark Leyland, University of Leicester.

^{15}N labelled C368 (Mr 7127 Da) was successfully expressed and purified to 99% purity (Figure 6.5.1). As expected, its NMR spectrum showed that C368 is predominantly unstructured. The spectrum was assigned through collaboration with Dr Ben Goult at the University of Liverpool. Aliquots of unlabelled SAP97 PDZ2 were then titrated into ^{15}N labelled C368. The chemical shift changes observed are shown in the overlay of titrated C368 in complex with PDZ2 with free C368, Figure 6.5.2. This data showed a disappearance of three of the C368 peaks indicating that they were involved in binding PDZ2. Thus, from this data we are able to

definitively define that residues E425, E427 and I428 of Kir2.1 form the binding site to SAP97 PDZ2. Based on chemical shift changes, it would appear that these three residues form the primary PDZ-binding site for Kir2.1.

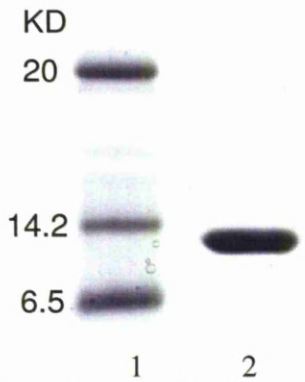


Figure 6.5.1. SDS PAGE gel showing concentrated C368 after Nickel affinity and anion exchange chromatography. Lane 1 = molecular weight markers Lane 2= C368.

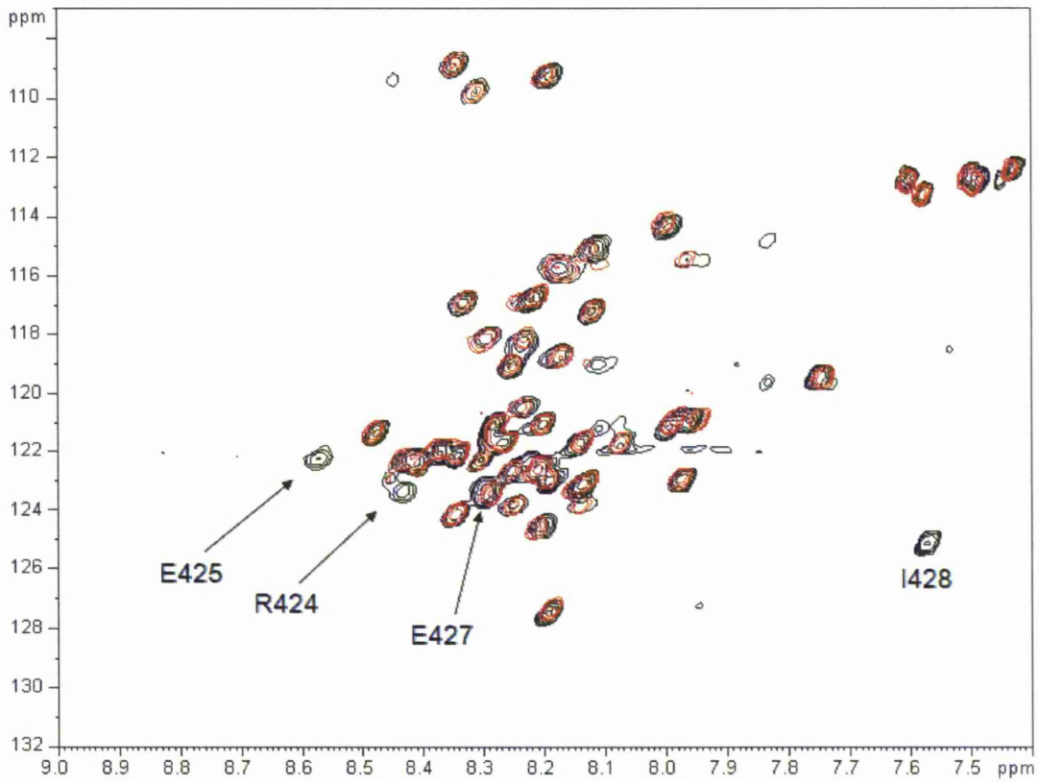


Figure 6.5.2. C368 : PDZ2 NMR spectrum. 15N-1H HSQC spectrum of C368 in complex with SAP97 PDZ2 domain at 298K. Black =free protein; Red =complexed protein 40 mM phosphate pH 6.8. The S426 peak at 8.07970, 115.14680 ppm in uncomplexed C368 is present but at very low intensity in the spectrum of the complex. Assigned residues are visible in free form but masked in complex.

6.6 Discussion

The SAP97 multi domain constructs each presented unexpected results during their characterisation. Δ546 consists of the U4-SH3-GK domains. The NMR spectrum of Δ546 consists of resonances from folded and unfolded protein domains, as concluded from the line widths and chemical shift characteristics of the resonances. For this protein the resonances from the folded region are very weak compared with the peaks from the unfolded region. This significant difference in line widths virtually led to most of the signals from the folded regions being masked by the intense resonances of the unfolded Hook region and prevented complete analysis of the data from this protein. It also made comparisons with the spectrum from the single GK domain very difficult.

Δ461 is an N-terminal extension of Δ546 with an additional PDZ3 domain. The NMR spectra showed more well-dispersed peaks when compared to Δ546. Overlay of Δ461 with the single PDZ3 spectrum shows that the extra peaks belong to the PDZ3 domain. The line width difference between the resonances are not as pronounced for Δ461 as for Δ546, suggesting that the flexible regions such as the U4 and/or Hook regions in Δ461 may be more constrained than in Δ546. The NMR spectra also showed that the presence of the PDZ3 domain does not significantly affect the chemical shifts of the SH3-GK resonances. This suggests that PDZ3 does not interact tightly with SH3-GK domains. In addition, the line widths of the SH3 domain are much narrower that those from the Δ546 domain. It can be concluded that the PDZ3 domain within Δ461 is rather more mobile and tumbles at a different rate than the SH3-GK. The SEC-MALLS for both Δ546 and Δ461 showed similar

Rg values, 2.0 verses 2.1. This suggests that both constructs are similar in size. This is somewhat counter-intuitive when considering the NMR data and the postulated mobility of the PDZ3 domain. It must be remembered that the MALLS method is more suited for proteins >100kD and therefore will not be able to distinguish small differences in size (the Mr difference between Δ461 and Δ546 is about 10kD).

The Δ70 NMR spectra showed many well dispersed peaks that through overlays were shown to belong to the PDZ domains and the U1 / N domain. Very few of the peaks were shown to belong to the SH3-GK domain. This suggests that the PDZ domains are much more mobile than the SH3-GK domain. Interdomain mobilities has previously been shown to exist in SAP97 and hence the NMR results are not unexpected (Goult et al., 2007). What is new here is the presence of the SH3-GK domain in the construct that appears not to have significantly constrained the movement of the PDZ domains. The SEC-MALLS data supports this view with an Rg twice that of Δ546 and Δ461; had Δ70 been more compact the Rg would be similar to Δ546 and Δ461.

CaM was shown to bind to SAP97 through fluorescence measurements. Previously through SPR CaM was shown to bind to the Hook region on SAP97 (Paarmann et al., 2002). It was also shown that the SH3 domain was not required for binding but without, it the binding was decreased 3 fold. Another finding was that binding only occurred in a calcium dependant manner (Paarmann et al., 2008). Interestingly, when Tavares et al (2001) attempted binding experiments between CaM and PSD95 SH3-GK domains they were unsuccessful. They believed that the Hook region was not large enough to interact with the CaM protein without being sterically hindered

by the GK domain and for binding to occur the intramolecular interactions through the SH3 split domain would have to be disrupted. This is thought not to occur in CaM binding to SAP97 because the Hook region of SAP97 contains a further 35aa.

In these new fluorescence experiments, SAP97 Δ461 showed the lowest binding affinity of 27.2 ± 1.5 μM; this could possibly have been due to the mobile PDZ3 domain moving or rotating away to a position that unblocked or unhindered the Hook domain binding site. In Δ70 the binding affinity is greater at 1.01 ± 0.2 μM. This difference in affinity from Δ461 could be explained by a conformation change induced by the presence the additional PDZ1 and 2 domains affecting the location of the PDZ3 domain. It is conceivable that the PDZ3 domain is 'displaced' from obscuring the CaM-binding region of the Hook domain. A similar difference in affinity of CaM for Δ461 and Δ546 was also observed in a parallel study in the laboratory when using ITC (unpublished data). Thus the more native folding of the Δ70 may promote the accessibility of the Hook domain leading to the greater K_d than for Δ461.

The fluorescence experiments gave a weak binding affinity for CaM to SAP97, at an affinity of 1 μM it could be argued that it is not physiological relevant because the CaM has many other binding partners which it will bind more strongly and thus will out compete the SAP97. This may suggest that the SAP97:CaM complex requires another protein present to increase the binding affinity to SAP97 or that the fluorescence data grossly underestimates the binding affinity of CaM to SAP97.

SAP97 GK domain was unfortunately not stable and could not be concentrated above 1 mM. This precluded extensive characterisation of the domain. The most logical reason for the instability is that the GK domain requires the SH3 domain for its stability. The SH3-GK pairing contains a split domain with the final SH3 β-sheet being formed by a polypeptide sequence downstream of the GK polypeptide sequence (McGee et al., 2001; Tavares et al., 2001). This β-sheet may also stabilise the GK domain.

A construct of the SH3 domain was attempted by our collaborators and myself but the addition of the spilt β-sheet proved unattainable and with pressing time constraints, focus was placed elsewhere.

The SAP97 NPDZ1 domain showed that the U1 / N region was unstructured and flexible. This flexibility was also observed through overlay of the Δ70 spectra. The NMR studies here contradict a previously reported intramolecular interaction between the N domain and the SH3-Hook domain (Wu et al., 2000). However, the Wu et al study (2000) used full length SAP97 and determined that residues 1-104 were required for intramolecular interactions with the SH3 Hook domains. Although no specific binding studies between SAP97 NPDZ1 and Δ546 were attempted due to time constraints there were no observations from the Δ70 NMR spectra to support an intramolecular interaction involving the N domain. Δ70 contains the N domain; if it was interacting with the SH3 and/or the Hook domain this would be expected to be visible in the NMR spectra, because the bound N domain would be stabilised and would therefore not appear as mobile and flexible. Nevertheless, more careful studies need to be undertaken to be absolutely certain.

In previous work from this lab it was shown, by NMR, that the 10 a.a. peptide, Kir2.1 C418 – 428 interacted with each of the PDZ123 domains of SAP97.PDZ2 was also shown to have the greatest affinity compared with PDZ1 and PDZ3 (Goult et al., 2007). The binding data using PDZ2 showed that R_{424}, E_{425}, and I_{428} from Kir2.1 C368 form the major sites of interactions with PDZ2. There is indication that both E_{427} and S_{426} are also involved in the interaction; this makes sense since the serine residue is a consensus residue with a PDZ-binding motif. Further work to probe the interactions between PDZ2 and Kir2.1 would involve determining the high resolution structure of the complex either by NMR or X-ray crystallography.

CHAPTER 7

Small Angle X-ray Studies of SAP97

7.1 Introduction

Small angle X-ray scattering (SAXS) is a tool to study the shape and conformation of a molecule. Unlike the high resolution NMR and X-ray crystallographic techniques SAXS does not define the three dimensional structure at atomic resolution. However, in combination with the high resolution techniques, SAXS has the ability to study the organisation of domains of molecules and provide information about conformation and shape. This is particularly useful in the situation when the proteins are either too large or too flexible for studies using NMR or X-ray crystallography. SAXS also has the advantage of being a solution-based technique.

As already stated earlier, no high resolution structures of any full length MAGUKs have been resolved. However, low resolution negative staining electron microscopy was used on full length SAP97 and showed that it existed in two major forms as extended rods (65 %) and as c-shaped or ring like structures (35 %) (Nakagawa et al., 2004).

SAXS was previously also used to determine the conformation of PDZ12 and PDZ123 domains (Goult et al., 2007); these data showed that the U4 linker and PDZ3 domain were mobile although the dynamics were restricted, with PDZ3 located within 40 Å from PDZ12.

This work focuses on SAXS analysis of the three SAP97 constructs Δ546, Δ461 and Δ70 (see Chapter 6 for details of protein preparation and other characterisation) and

will attempt to elucidate their shape and the orientation of the individual SAP97 domains. The methods used involve many computational steps that are outlined in the flow diagram below, Figure 7.1.1. The individual programs used are discussed in chapter 2.

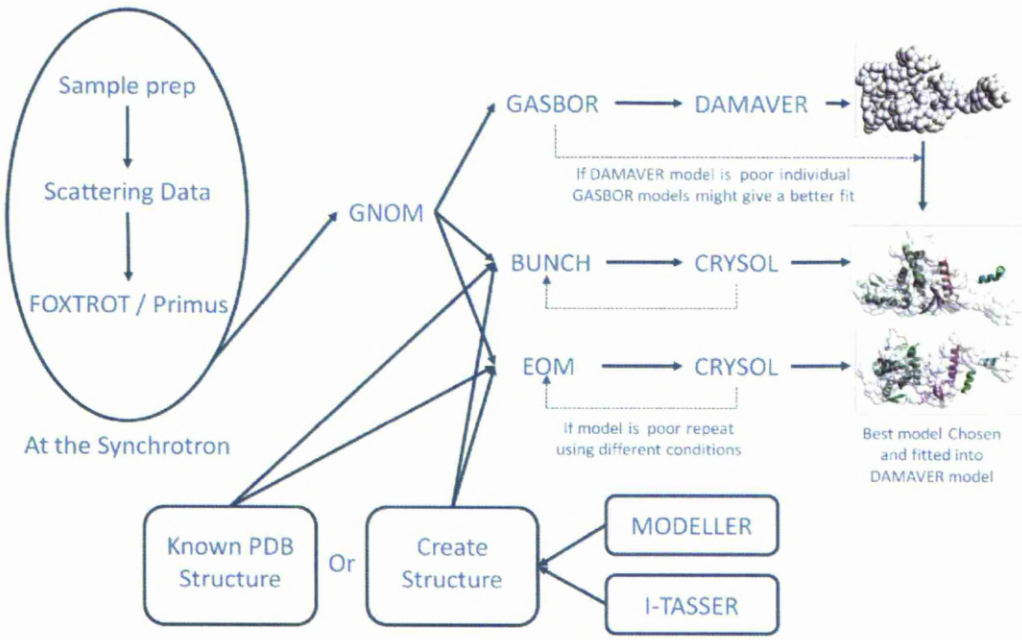


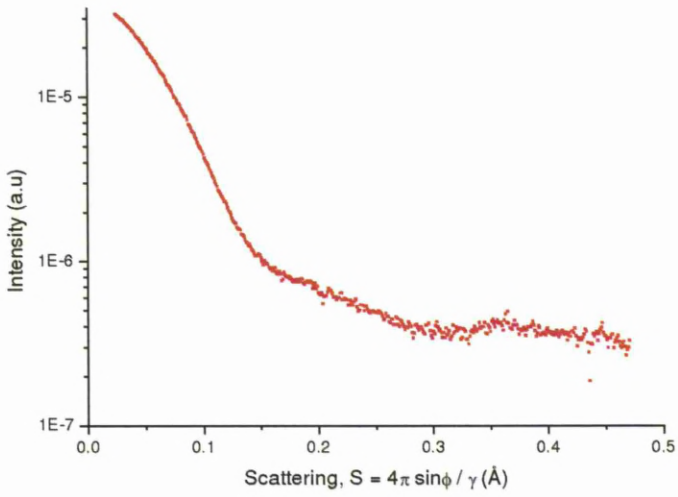
Figure 7.1.1. Flow diagram of SAXS interpretation. Initial data collection and data processing is carried out at the synchrotron or X-ray source. The remaining interpretation requires freeware programs and can be processed on any modern computer.

7.2 SAP97 Δ546 Results

7.2.1 SAP97 Δ546 Scattering Results

The SAXS data for three different concentrations, 1, 2, and 3 mg/ml was collected in batch mode on the SWING beamline at the SOLEIL Synchrotron, Paris. The data was initially analysed by the synchrotron principal beamline scientist, Dr Javier Pérez using the in-house software "Foxtrot" and freeware program PRIMUS which averaged data for the three concentrations and then subtracted the background scattering. The SAP97 Δ546 scattering data was then further processed using the program GNOM and obtained a D_{max} of 115 Å and an R_g 29.2 Å, distribution of scattering mass and radius of gyration respectively,(Figure 7.2.1) where the D_{max} is the maximum dimension of the proteins and the Rg is a measure of the size of the protein derived from the root mean square distance of the central particle to the scattering length density distribution.

The P(r) value (regularization parameter) was also calculated directly from scattering curve using GNOM (Figure 7.2.1) which depicts the distribution of intramolecular atomic distances within the molecule (Svergun, 1992). The broad peak indicates a spherical shape and the tail possibly indicates a smaller longer area present.



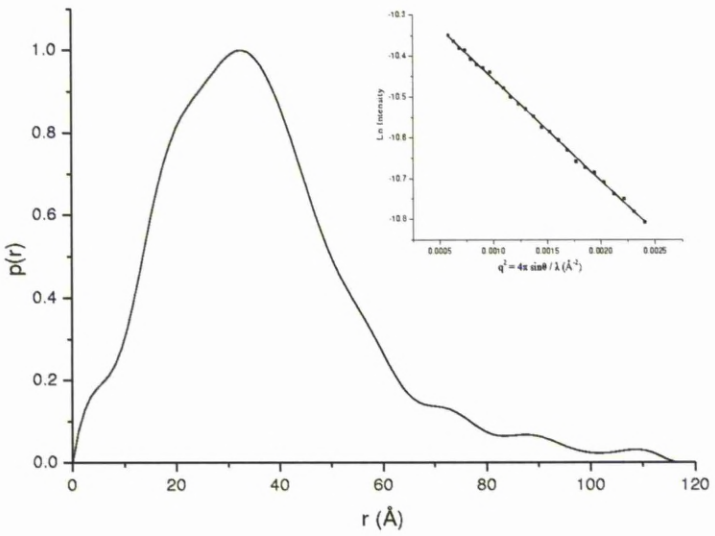


Figure 7.2.1. Small angle X-ray scattering of SAP97 Δ546. The averaged scattering data from 1, 2 and 3 mg/ml concentrations of SAP97 Δ546. The lower plot shows the P(r) value (regularization parameter) calculated by GNOM and gives the Rg 29.2 Å and the D_{max} 115 Å.

7.2.2 SAP97 Δ546 GASBOR and DAMAVER

Although the P(r) enables the molecular shape to be deduced, there are now algorithms that allow the production of 3D models from the 1D scattering data (Svergun and Koch, 2002). GASBOR is an *ab initio* modelling program that uses dummy residues (small spheres) to build a compatible model inside the D_{max} of the molecule that allows the fitting of the 1D scattering curve up to 0.5 nm resolution (Svergun et al., 2001). Models of Δ546, were created using GASBOR and were then averaged using the program DAMAVER (Volkov and Svergun, 2003).

The Δ546 models or envelopes generated by GASBOR (n = 10 no rejections) each showed a large compact area with lumps and protrusions that in some cases showed a hollow centre with a second smaller elongated area protruding at one end (Figure 7.2.2). It is most likely that the protrusion is the N-terminal U4 linker and hexahistag tail of Δ546 (see Appendix for the complete polypeptide sequence of Δ546). When averaged by DAMAVER (Figure 7.2.3) the envelope created showed a smooth large spherical area that has a depression on one side followed by a second smaller sphere protruding at one end.

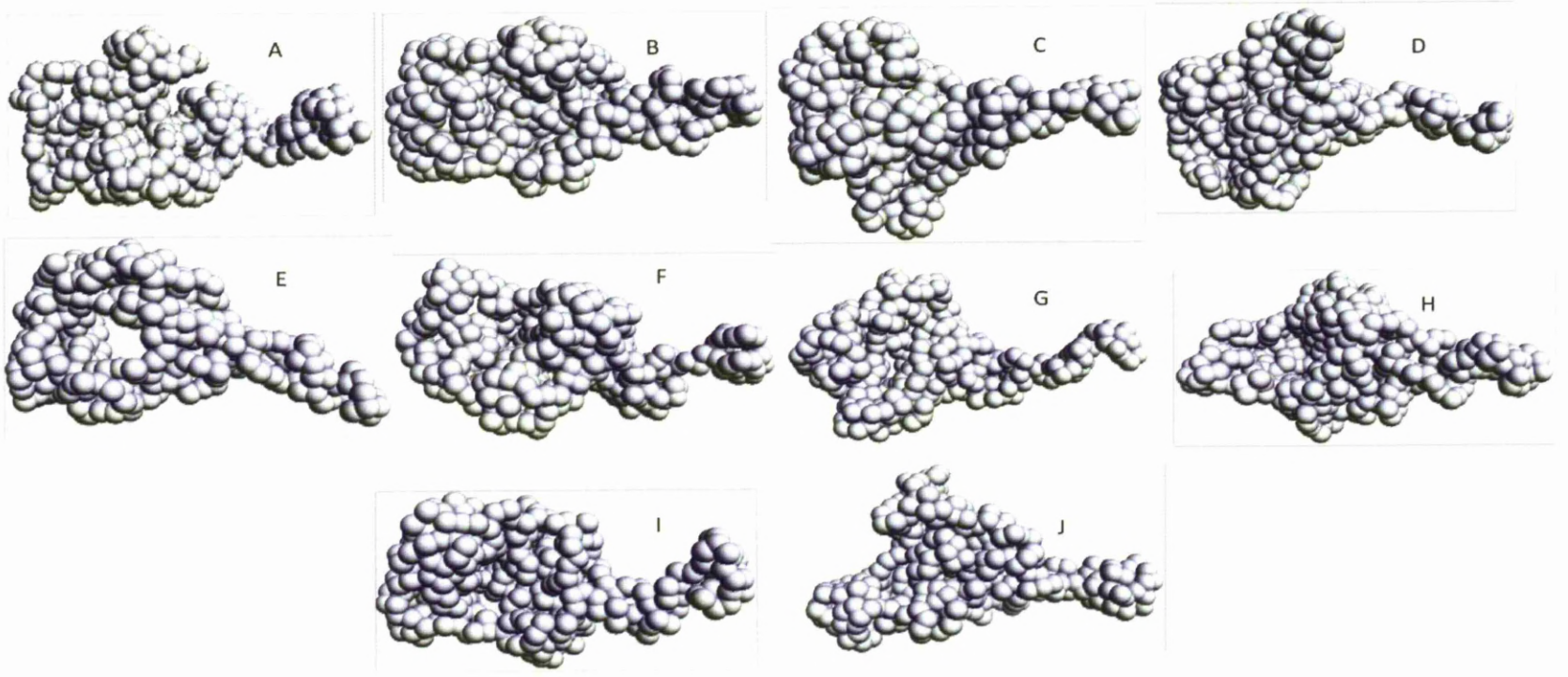


Figure 7.2.2. Individual model envelopes for SAP97 Δ546. 3D envelopes generated from 1D scattering data using program GASBOR, A-J show the 10 replicates generated from 10 individual runs. Each model is approximately 90 Å in length.

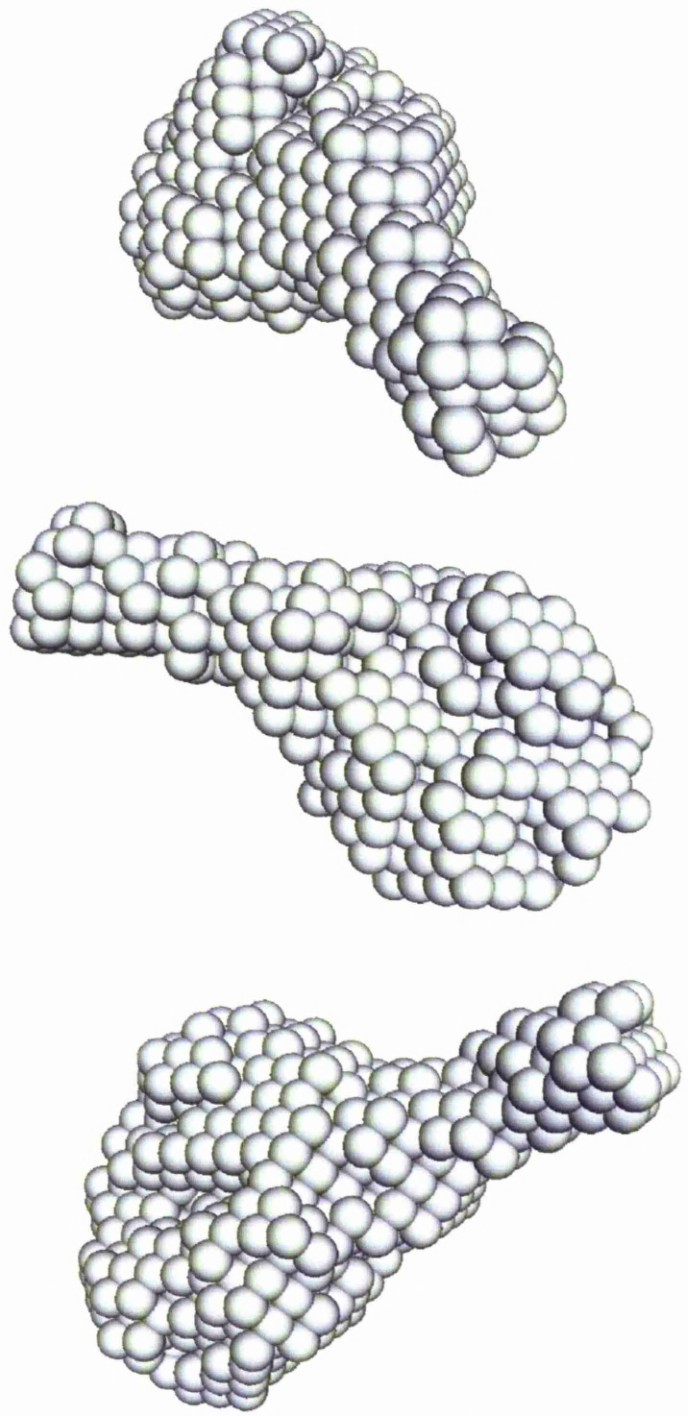


Figure 7.2.3. Averaged model envelope for SAP97 Δ546. Three orientations of the 3D envelope generated from 1D scattering data using program GASBOR and averaged using program DAMAVER (n=10).

7.2.3 SAP97 Δ546 Rigid-Body Modelling

Multidomain proteins can be fitted to the scattering data using rigid body modelling programs such as BUNCH or EOM. This requires high resolution structures for the individual domains form either NMR or X-ray crystallography. This provided problems for SAP97 that has no high resolution structures available for the SH3 and GK domains, although X-ray and/or NMR structures for the individual PDZ domains do exist. Fortunately there are structures available for PSD95 whose SH3-GK domain shares 70 % homology with SAP97, and where the main difference between PSD95 and SAP97 is in the Hook region. In SAP97 this region, also known as the U5 linker, is elongated (80 amino acids in SAP97 rather than 35 in PSD95). This allows PSD95 to be used as a template to create models of SAP97 SH3 and GK domains using *ab initio* or comparative modelling using, respectively, I-TASSER and MODELLER programmes.

7.2.3.1 I-TASSER Modelling

A model of SAP97 SH3-GK domain was generated using the *ab initio* modelling program I-TASSER (discussed in chapter 2). The SAP97 Δ546 sequence was input into the web server and 5 models generated using the PSD95 (1KJW) and other SH3 and GK domains from the CaV β-subunit-2a (1T3L) and a PDZ domain (2FY5) as templates. To test how reliable the models were the program CRYSOL (Svergun et al., 1995) was used that fit the model to the scattering data (Figure 7.2.4). The best fit was given by model 5 so this model was used for the further experiments. Only the individual domains were required for the rigid body

modelling because the programs would add any unstructured areas or linker regions. The coordinates of only the SH3 (SH3mod5), GK (GKmod5) domains and two separate helices that formed the U4 linker were selected from the PDB files of the models to create the templates for the rigid body modelling (Figure 7.2.5).

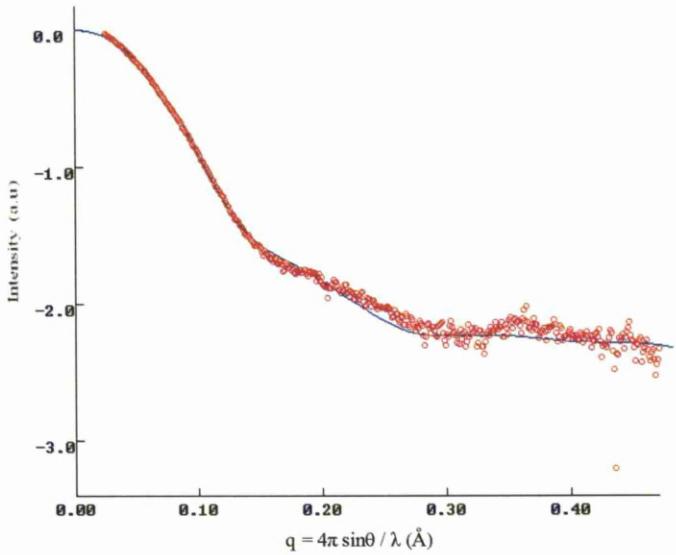


Figure 7.2.4. CRYSOL curve for I-TASSER model of SAP97 Δ546. The curve fit in blue represents how close the I-TASSER model generated is to the experimental scattering data (in red).

The SH3mod5 and GKmod5 domains were overlaid with the PSD95 SH3 and GK domains and showed good chain alignment; however, not all of the β-sheets of the PSD95 model were replicated in both SH3mod5 and GKmod5 (Figure 7.2.6). To improve the models individual SAP97 SH3 and GK domains were then generated with I-TASSER using specific domain sequences without the linker regions. I-TASSER only predicted one result for both domains, SH3Tass1 and GKTass1. These new models when overlaid with PSD95 showed a greater similarity and showed a far better alignment of the β-sheets (Figure 7.2.7).

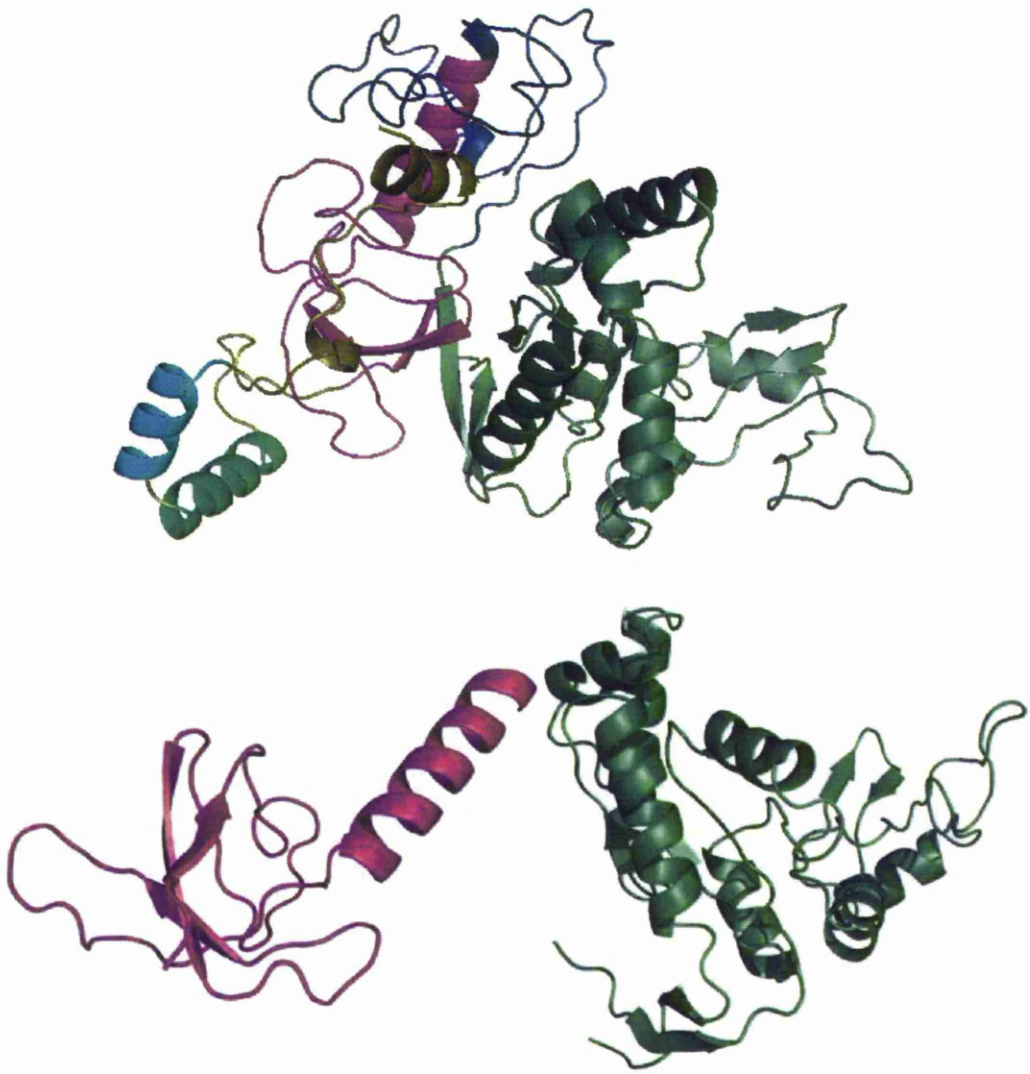


Figure 7.2.5. I-TASSER model of SAP97 Δ546. The amino acid sequence of SAP97 Δ546 was inserted into I-TASSER on line program and 5 models were generated. (Top) The model that gave the closest match to the scattering data as judged by CRYSOL. The individual domains have been coloured, SH3 Magenta, GK Forrest green, Hook domain blue and the U4 linker α-Helices in green and cyan. (bottom)The SH3 (SH3mod5) and GK (GKmod5) domains 'excised' from the above model, leaving out the flexible linker regions.

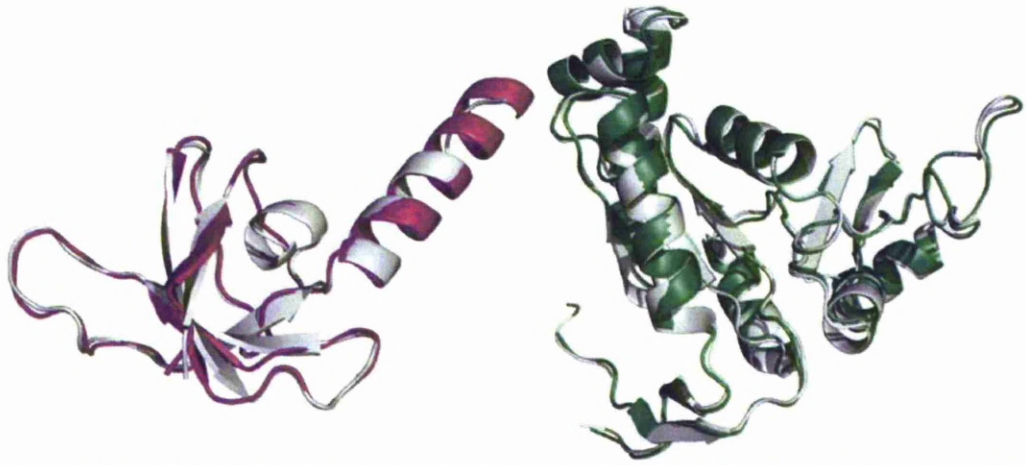


Figure 7.2.6. PSD95 SH3, GK domains overlaid with SAP97 SH3mod5 and GKmod5 models generated using I-TASSER. SH3 Magenta, GK Forrest green. The PSD95 (grey) was taken from the crystal structure 1KJW (McGee et al., 2001). PyMOL was used for the overlay.

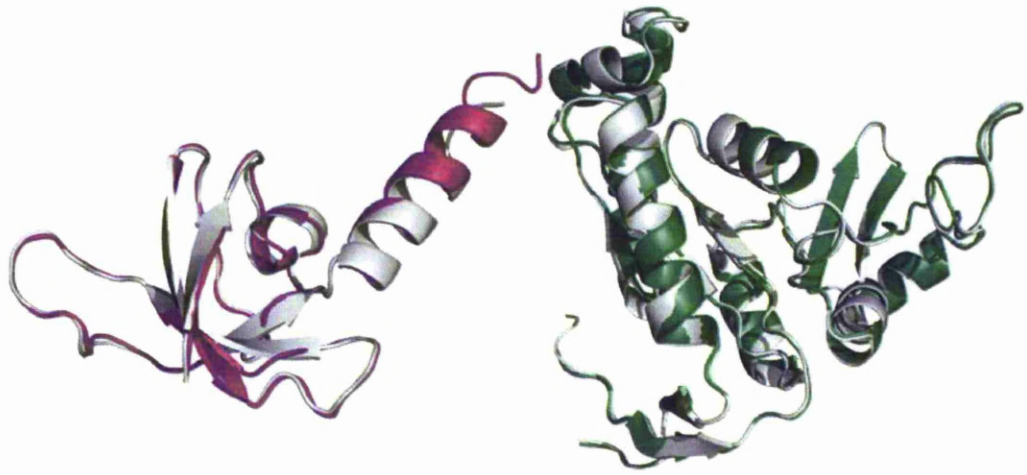


Figure 7.2.7. PSD95 SH3, GK domains overlaid with SAP97 SH3Tass1 and GKTass1 generated models. Models generated using I-TASSER on individual domains of SAP97, SH3 Magenta, GK Forrest green. The PSD95 (grey) was taken from the crystal structure 1KJW (McGee et al., 2001). PyMOL was used for the overlay.

7.2.3.2 MODELLER Modelling

In comparative modelling, the SAP97 SH3 and GK domains models were also generated using MODELLER with a PSD95 template. There are two crystal structures for PSD95 in the PDB database (1KJW and 1JXO, respectively, from McGee et al. (2001) and Tavares et al. (2001).). The SH3 domains of both are different with the α-Helix longer in 1JXO than 1KJW. The 1KJW pdb file shows all the α-helices, β-sheets and residues in between intact whereas 1JXO pdb file has one of the loops of residues undefined. The GK domains do not differ noticeably. The template 1KJW was used because of the intact loop of residues that may have proven more difficult for MODELLER to predict correctly.

The 5 models predicted for both SAP97 SH3 and GK domains were analysed by DOPE, an analysis program used with MODELLER to show goodness of models pictorially. For SAP97 SH3 the plotted DOPE results (Figure 7.2.8) showed a good comparison between each of the MODELLER predicted models with the exception of the second model, SAP97SH32 that varied in the last 10 residues. The overlay of all the MODELLER predictions shows that there is much variation at the C-terminus of the SH3 domain as this is the least constrained region. Also included in the DOPE analysis was the SH3mod5 model from I-TASSER (named model5C and model5N in Figure 7.28); this too matched up well except for the last10 residues. This is because it was modelled using both PSD95 templates (1JXO and 1KJW) and this facilitates a greater change in the α-Helix motif.

The SAP97 GK predicted domains all showed similar DOPE profiles with the 1KJW template except for the first 10 residues and the last 5 residues. When

looking at the overlaid models both the C- and N-terminus of the domains are unstructured, which may account for the variation observed in the DOPE analysis.

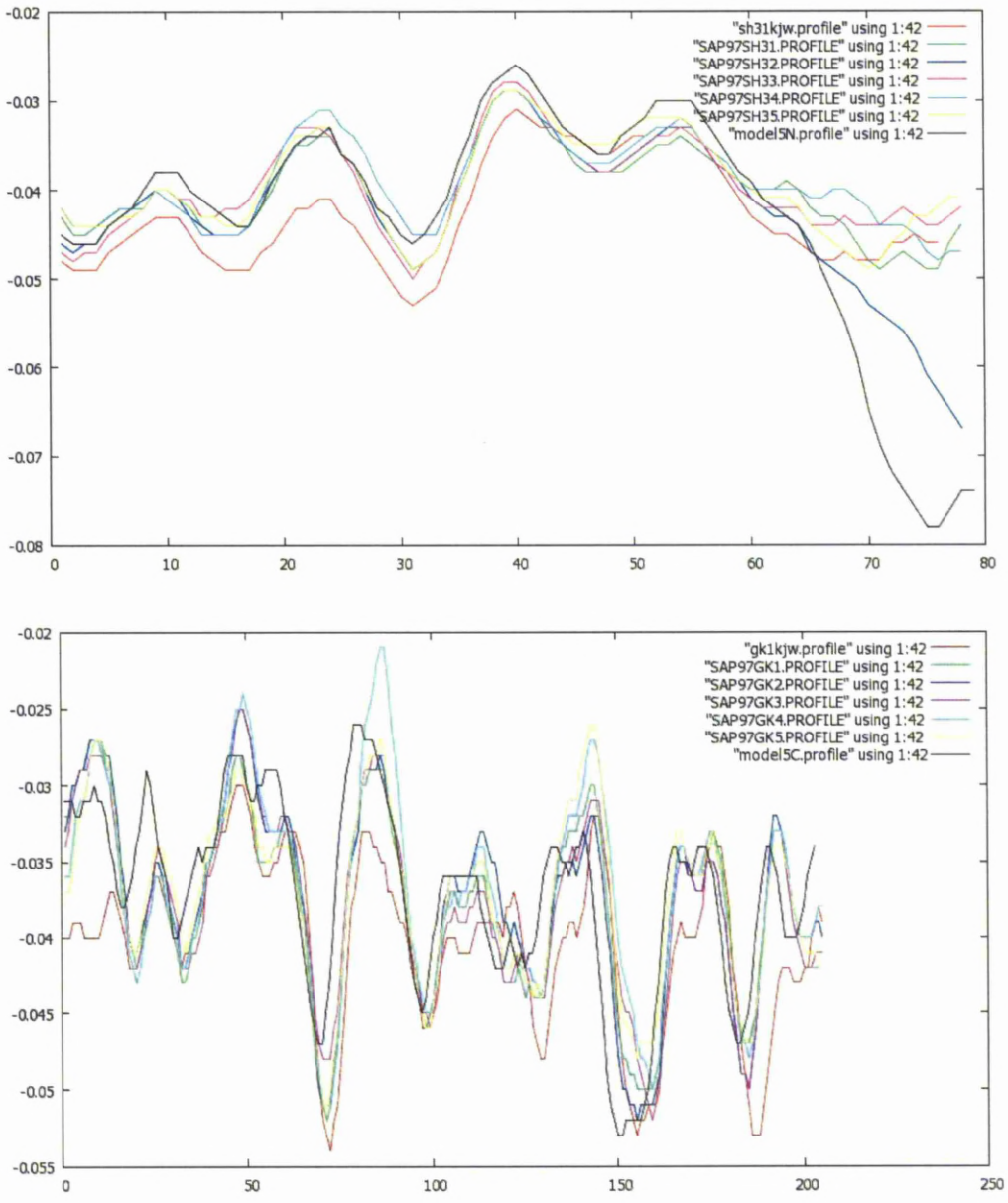


Figure 7.2.8. DOPE analysis for SAP97 SH3 and GK domains. MODELLER generated DOPE curves showing the energy differences between the PSD95 (1KJW) template and 5 generated profiles, top SH3, bottom GK domain.

Figure 7.2.9. SAP97 SH3 GK domains generated by MODELLER overlaid with PSD95. The 5 models generated show good comparison with PSD-95. Same colour scheme used as Figure 7.2.8 DOPE analysis. The PSD95 (grey) was taken from the crystal structure 1KJW (McGee et al., 2001). PyMOL was used for the overlay.

7.2.3.3 I-TASSER – MODELLER Comparison

Models of the SAP97 SH3 and GK domains generated with I-TASSER and MODELLER were evaluated using the RMSD when overlaid against the PSD95 template, 1KJW in PyMOL (Table 7.2). CRYSOL could not be used to test reliability. CRYSOL measures the model against the reference scattering data and because each of these models was of individual domains only, they would not fit the scattering data. The RMSD results show that MODELLER generated structures for both SAP97 SH3 and GK domains gave lower RSMD scores and these are interpreted as better models. There are two possible reasons for this:

(a) the MODELLER program concentrates not only on main chain orientation but also on the side chains of the model when compared with I-TASSER. This can clearly be seen when SAP97SH32 (Modeller) and SH3Mod5 (I-TASSER) are overlaid with 1KJW (Figure 7.2.10). Whether or not the side chains predicted

positions from MODELLER are actually correct for SAP97 cannot be determined until higher resolution models of SAP97 are solved.

(b) The I-TASSER models use a combination of both of the PSD-95 crystal structures, 1KJW and 1JXO so when compared to just the 1KJW structure they show a greater RSMD.

Table 7.2.SAP97 SH3 and GK RMSD comparisons of I-TASSER and MODELLER models using PSD95 1KJW as a template.

SH3 Model	RMSD (Å)
SH3Mod5 (I- TASSER)	1.03
SH3Tass1 (I-TASSER)	0.807
SAP97SH32 (Modeller)	0.422

GK Model	RMSD (Å)
GKMod5 (I- TASSER)	0.952
GKTass1 (I-TASSER)	0.696
SAP97GK1 (Modeller)	0.322

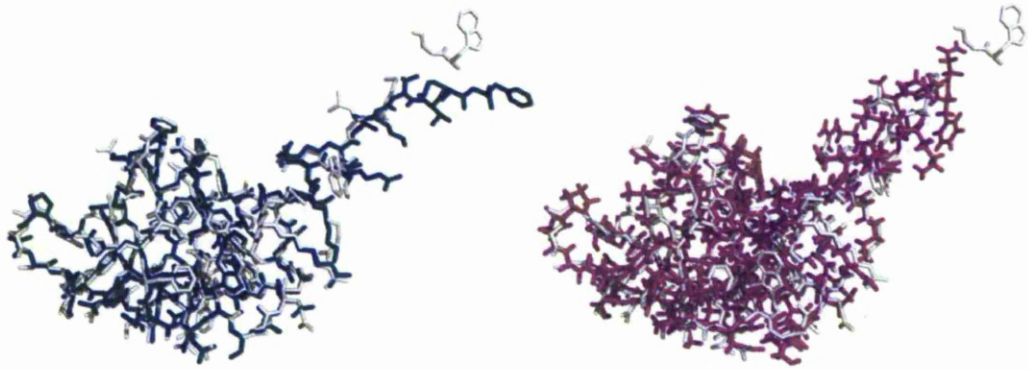


Figure 7.2.10. SAP97 SH3 domain side chain overlays. The MODELLER generated SAP97SH32 (left) and I-TASSER generated SH3Tass1 (right) models showing only backbone and side chains overlay with PSD95 SH3. The PSD95 side chains (red) was taken from the crystal structure 1KJW (McGee et al. 2001). PyMOL was used for the overlay.

The choice of which SAP97 SH3 and GK models to use for the rigid body modelling really comes down to personal preference as all models are acceptable. Since the SAXS scattering can only interpret globular proteins, the finer twists and turns of α-helices and β-sheets or position of side chains will not influence the modelling. For this reason both models generated from I-TASSER and MODELLER will be used.

7.2.3.4 SAP97 Δ546 Bunch Rigid body modelling

The SH3Mod5 and GKMod5 SAP97 models and the two α-Helices that constituted the U4 linker were fitted to the scattering data using the program BUNCH (See chapter 2 for explanation). The SH3Mod5 and GKMod5 were fixed to show the *ab initio* modelling and the possible positions of the U4 linker and the Hook region (Figure 7.2.11). When six BUNCH runs were overlaid the U4 linker region, that contained the two short helices, was seen to be very flexible and able to exist in many positions that surround the core SH3-GK domain and are often elongated. The Hook was modelled and shown to lie between the SH3 and GK domains.

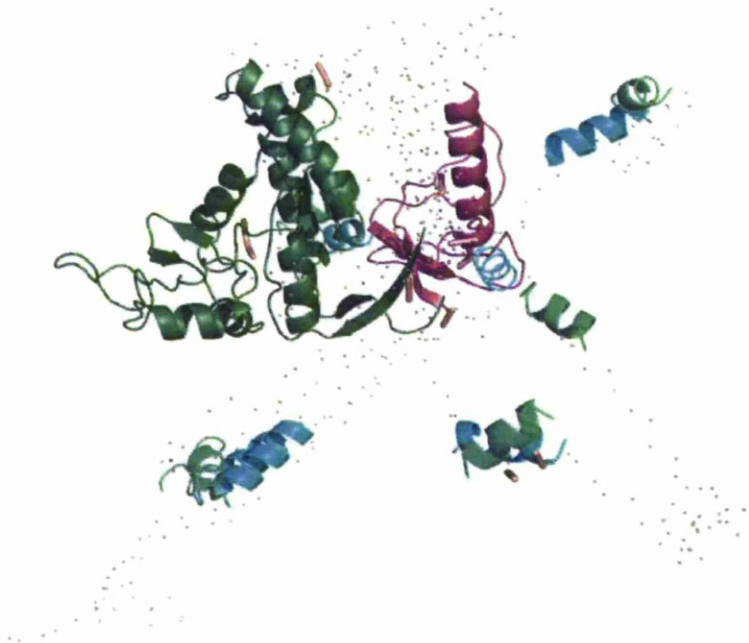


Figure 7.2.11. SAP97 Δ546 Rigid Body modelling. An overlay of 6 BUNCH analyses with GK (forest green) and SH3 (magenta) fixed and two α-Helices that form part of the U4 linker (cyan, green) flexible. The individual dots each represent the linker region or Hook region

To test which of the BUNCH models best fit to the scattering data the .pdb files of each model was passed through CRYSOL. 546r2, the second BUNCH run, gave χ = 11.2 which is well within tolerances for multi domain proteins (Petoukhov and Svergun, 2005). The BUNCH model also showed an intact modelled Hook region and a more compact U4 linker region (Figure 7.2.12).

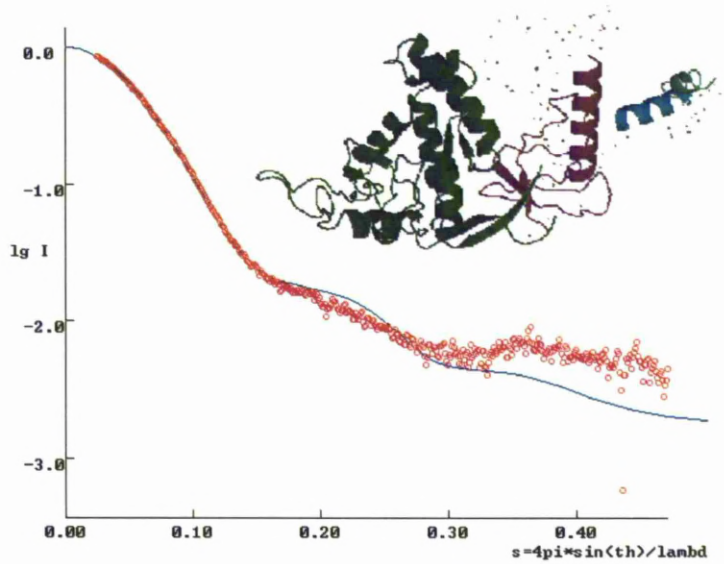


Figure 7.2.12. Analysis of SAP97 Δ546 by BUNCH and CRYSOL. Shown in the inset is the Δ546 conformer analysed for goodness of fit by CRYSOL

The BUNCH model was then superimposed over the averaged GASBOR image (Figure 7.2.13) and visually showed a poor fit. There are large areas of free space around both the GK and SH3 domains and the base of the SH3 domain does not fit very well within the envelope. The Hook region could possibly flex to fill some of the spaces but not all. The U4 linker region also does not fit the envelope but this is of less concern as the flexibility in the unstructured region could move within the envelope.

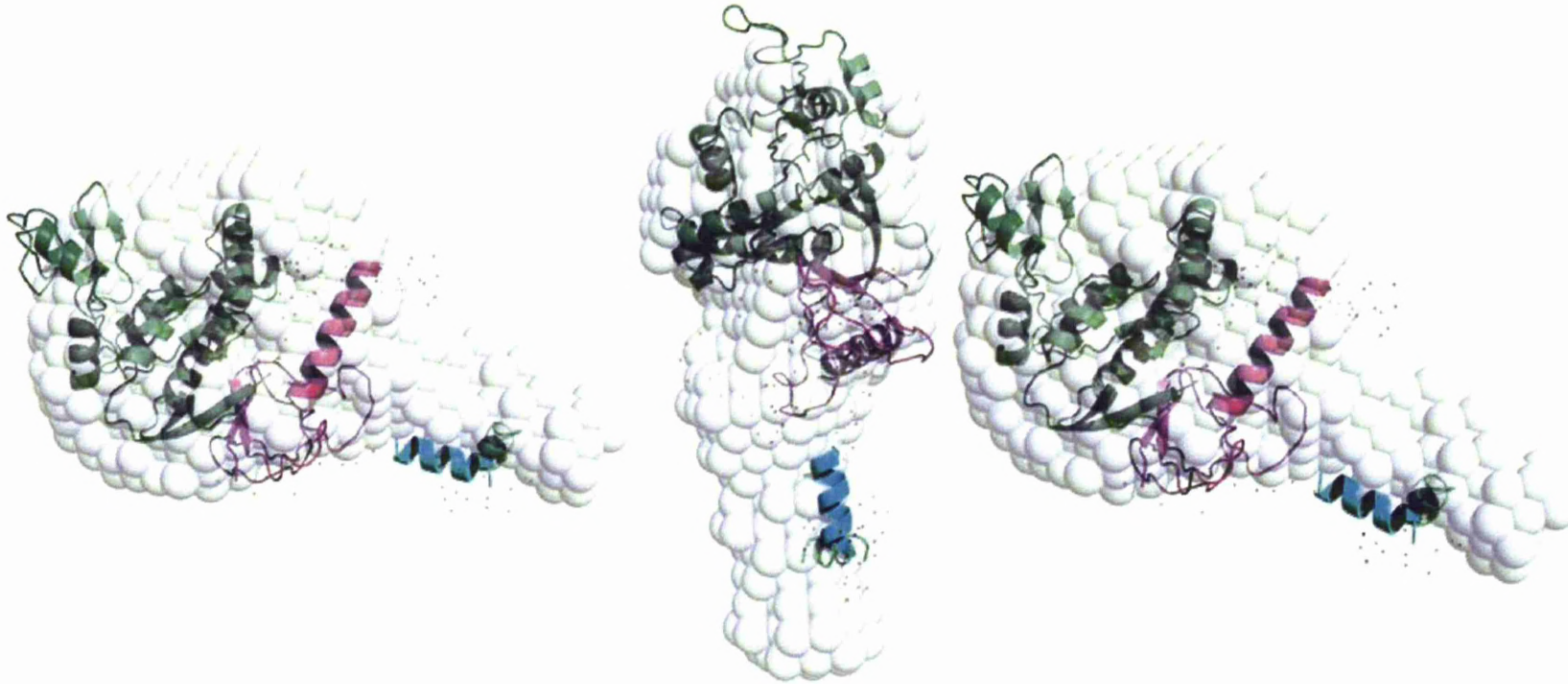


Figure 7.2.13. Shape reconstruction of SAP97 Δ546 from experimental scattering data. The averaged GASBOR envelope is rotated with the SAP97 Δ546 BUNCH model 546r2 superimposed on top. GK (forest green) and SH3 (magenta) fixed and two α-Helices of the U4 linker (cyan, green).

7.2.3.5 SAP97 Δ546 EOM Rigid Body Modelling

To attempt to improve the superimposed fit the SAP97 SH3-GK domains were analysed using Ensemble Optimisation Method, EOM, another rigid body modeller that was developed for proteins which are more unstructured or flexible in nature (Bernado et al., 2007). The expectation was that EOM might help to determine the position of the Hook region. Initially SH3Mod5 and GKMod5 models were used for consistency (section 7.2.3.4) however, this resulted in the two domains being separated which is physiologically impossible because of the split domain that the GK and SH3 domains share. To remedy this, the BUNCH 546r2 model was used but the flexible U4 region was removed and replaced with an U4 alpha helix from the PDZ3 domain (1PDR). EOM created 10000 theoretical curves and then selected the ensemble of curves that best fitted the experimental data. Of these, the conformers that showed the greatest frequency or abundance were selected and models generated.

The SAP97 Δ546 EOM frequency distribution (Figure 7.2.14) shows the pool of 10000 structures spread from 24 to 34 Å while the Δ546 conformers best fitting the data is spread from 24 to 29 Å with an average R_g ~25 Å. This concurs well with an R_g ~29.2 Å that was calculated by GNOM (see section 7.2.1) This shows that Δ546 exists in a predominantly compact form with 60 % of conformers between 24 to 29 Å. The more elongated conformers are most likely caused by elongation of the U4 linker.

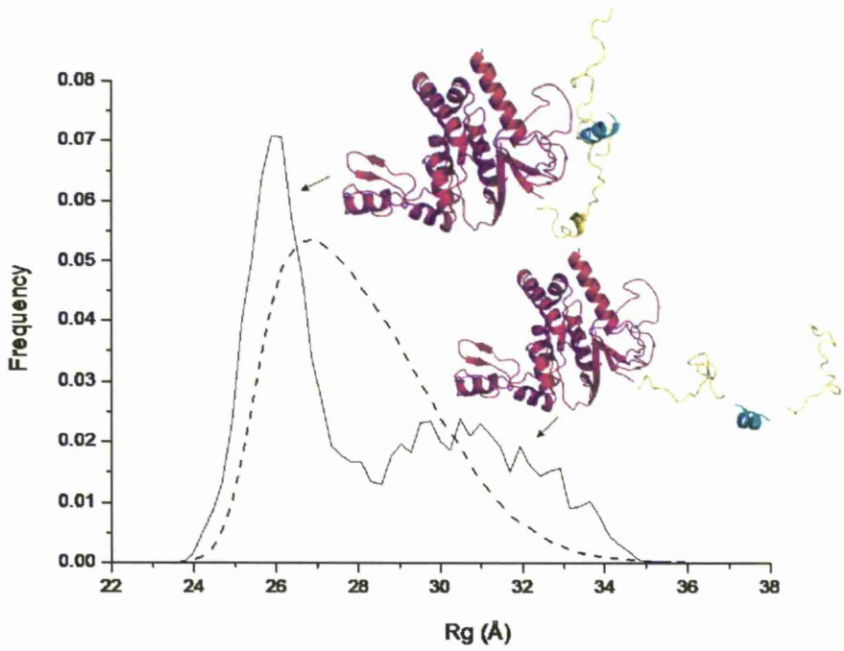


Figure 7.2.14. Analysis of SAP97 Δ546 by EOM. A frequency distribution of the SAP97 Δ546 conformers as a function of R_g. Black dashed line is the distribution of the initial 10000 pool conformers and the black line is the distribution of the analysed Δ546 ensemble. Two representative structures for the conformers within the EOM ensemble were judged to realistically fit the Δ546 model

In total, the ensemble generated 17 conformers where the U4 linkers were either close to the SH3 domain or completely extended giving a very elongated model.

When the intensities plotted as a function of momentum of transfer, a measure of fit equivalent to CRYSOL (Figure 7.2.15) it was shown to give a good fit, χ^2 4.79 that was lower than the fit given by BUNCH.

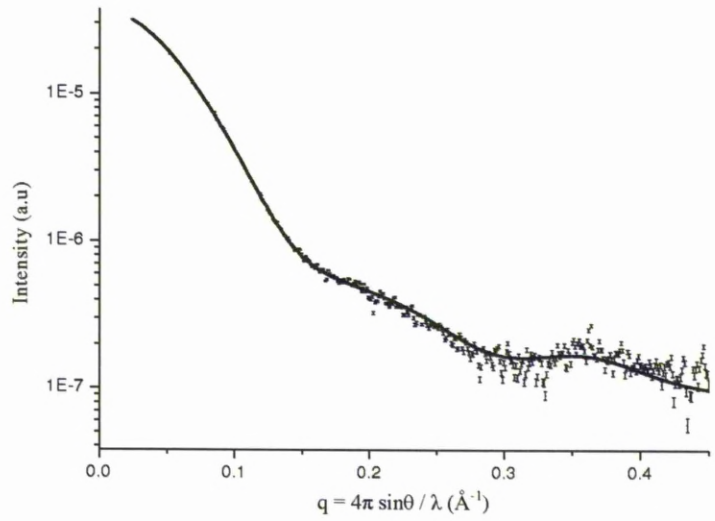


Figure 7.2.15. Analysis of SAP97 Δ546 by EOM fitting. SAXS intensities plotted as a function of momentum of transfer. Dots represent experimental data; solid line is fit obtained from the ensemble of 14 conformers

The 17 EOM models were superimposed over the SH3-GK domain and overlaid on one GASBOR model to show the conformational variety for U4 region which contains the α helix (Figure 7.2.16). The U4 linker can be seen to vary in position that suggests that it is highly mobile.

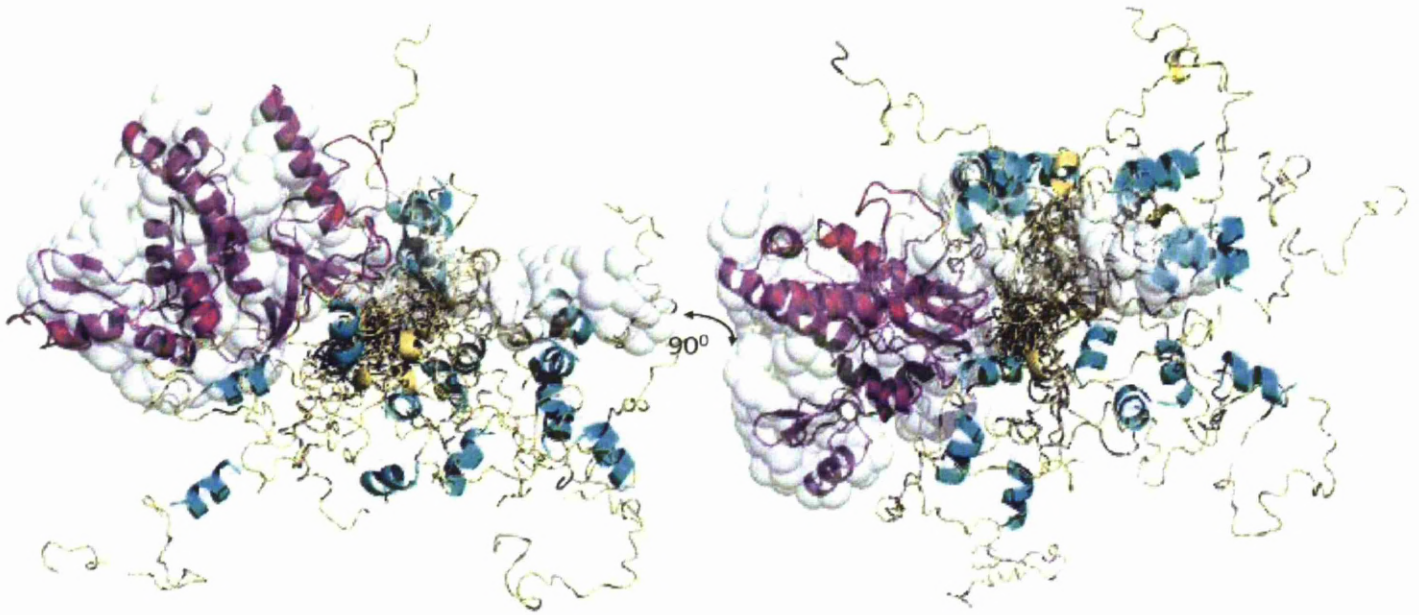


Figure 7.2.16. Overlay of SAP97 Δ546 EOM models. The 17 Δ546 EOM conformers were overlaid to view the differences in each conformer. The right view is a 90° rotation. The 546r2 model (magenta) was fixed and α-Helices of the U4 linker (cyan,) and the linker region (gold) are fit onto an individual GASBOR shape.

7.2.3.6 SAP97 Δ546 Bunch and EOM-GASBOR Overlay

When the SAP97 Δ546 BUNCH model was overlaid with the averaged GASBOR model it was not a good fit, with areas of the envelope volumes unoccupied. The SAP97 Δ546 EOM models were also overlaid but these too did not create a good fit. When analysing the individual GASBOR envelopes (Figure 7.2.2) it was noticed that some envelopes provided a smaller volume that might fit the rigid body models more tightly. The BUNCH and EOM models were then superimposed over the individual envelopes. SAP97 Δ546 BUNCH model 546r2 fit well into the envelope J; for comparison PSD95 crystal structure was also successfully fitted into the same envelope. The EOM models also superimposed well into the individual envelopes with SAP97 Δ546 EOM 3588 fitting into the GASBOR G individual model respectively (Figure 7.2.17 a-c). The linker regions in each overlay did not always match up with the envelope that is most likely caused by the flexible nature of the linker region.

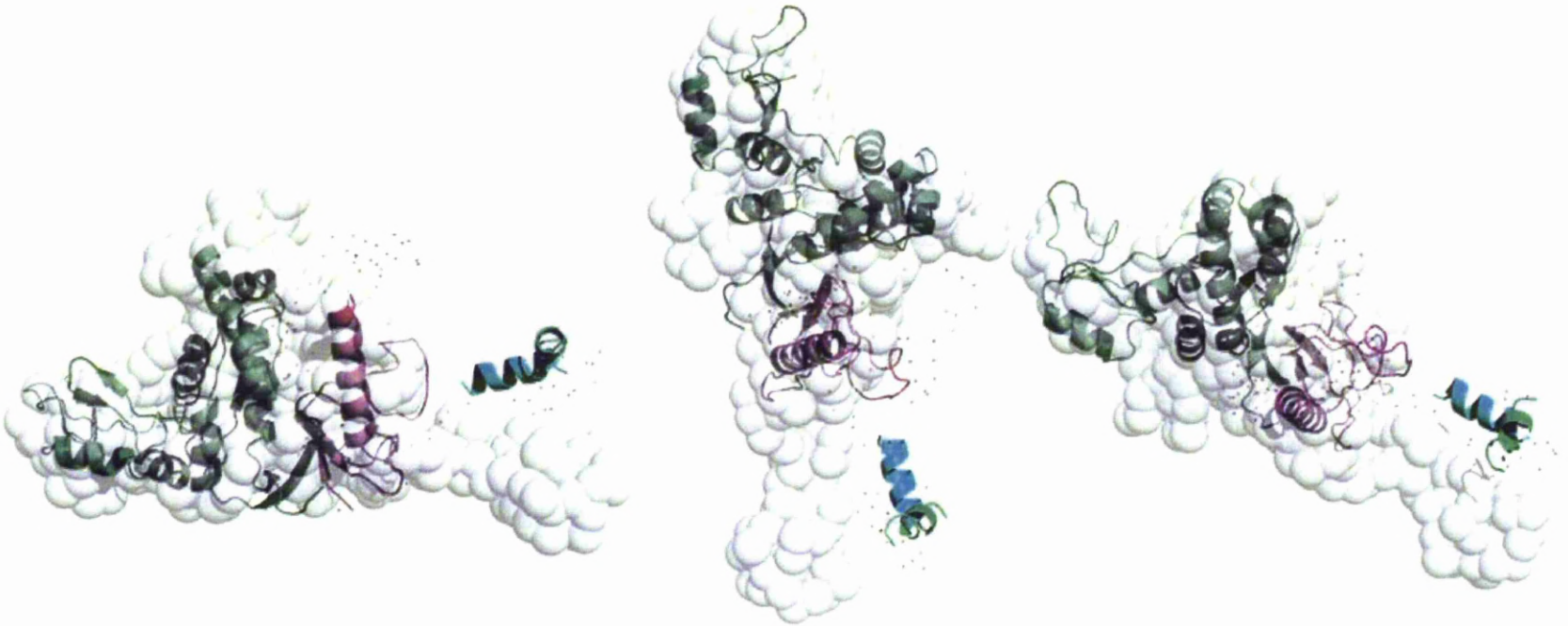


Figure 7.2.17a. Shape reconstruction of SAP97 Δ546 from experimental scattering data. GASBOR envelope J is rotated with the SAP97 Δ546 BUNCH model 546r2 superimposed on top. GK (forest green) and SH3 (magenta) fixed and two α-Helices of the U4 linker (cyan, green).

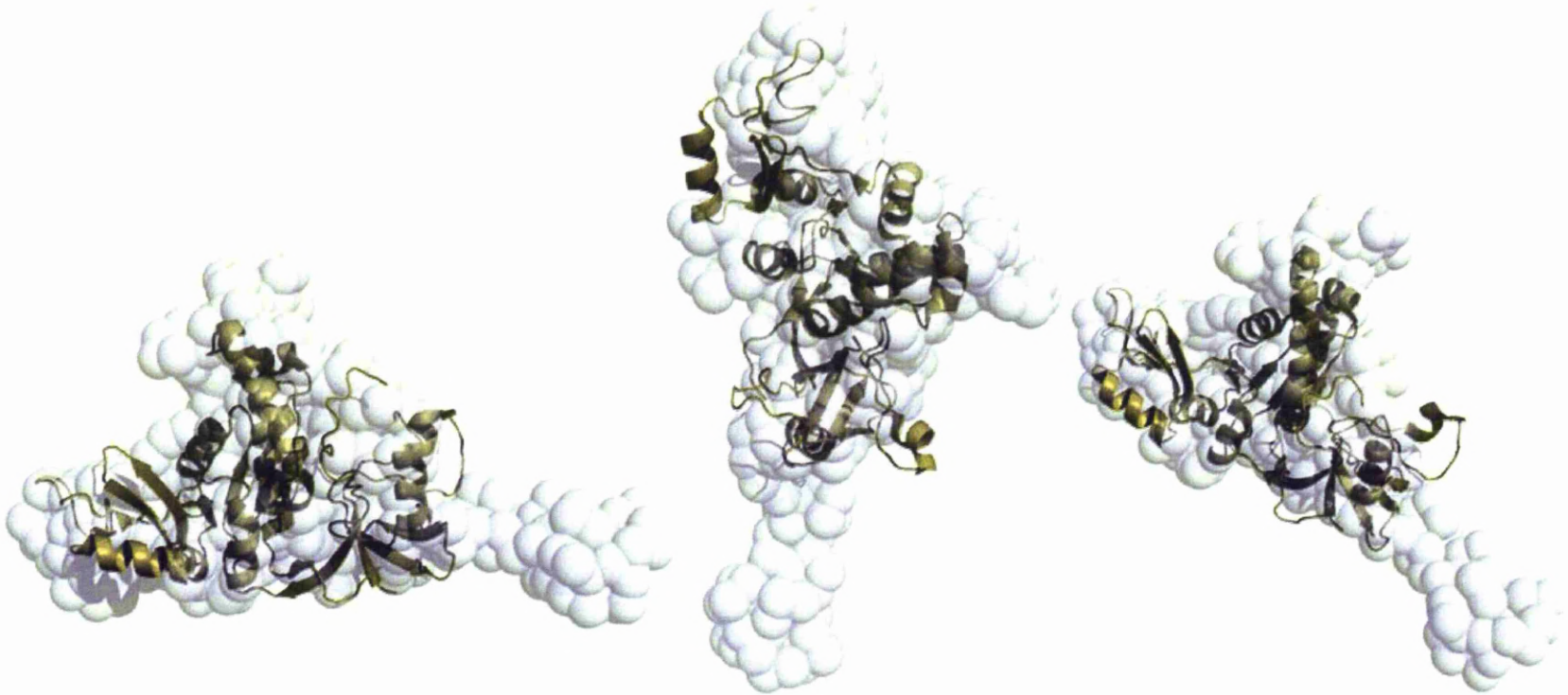


Figure 7.2.17b. Shape reconstruction of high resolution structure PSD95 and GASBOR envelope from experimental scattering data. GASBOR envelope J is rotated with the PSD95 1JXO superimposed on top (Orange).

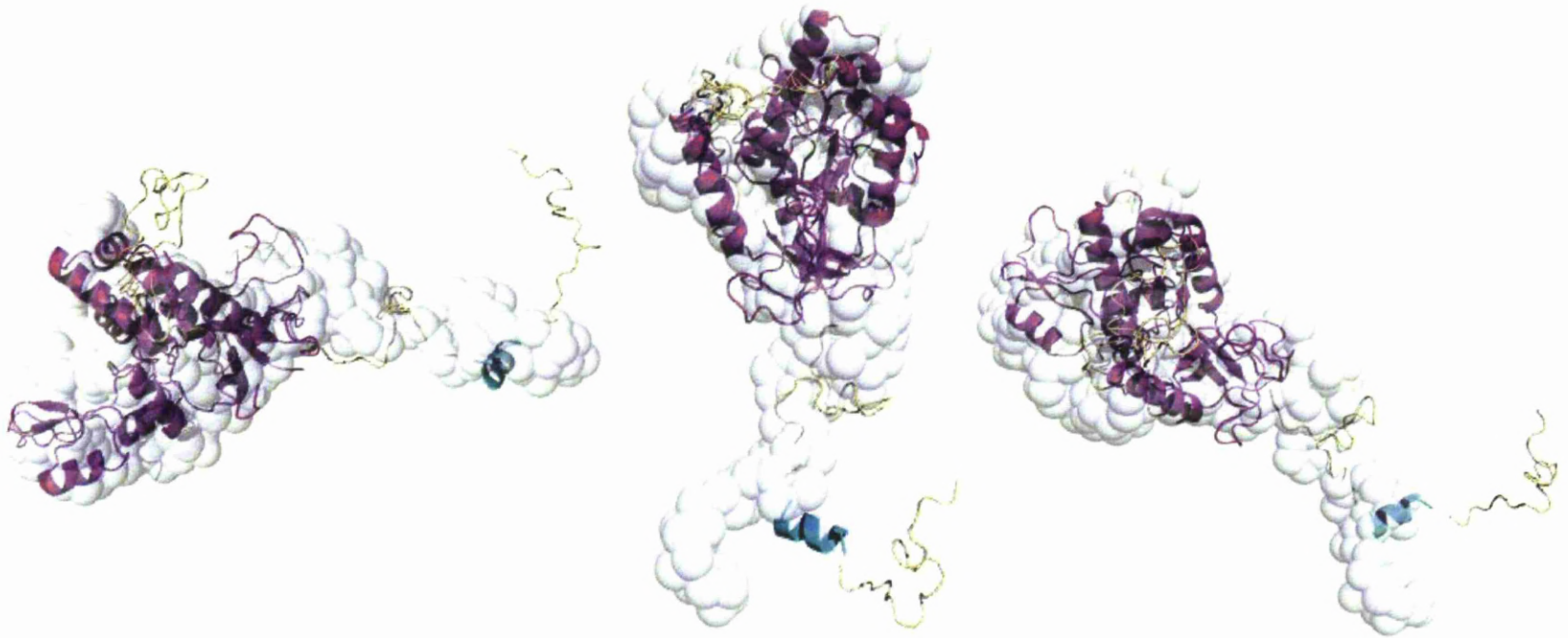


Figure 7.2.17c. Shape reconstruction of SAP97 Δ546 from experimental scattering data. GASBOR envelope G is rotated with the SAP97 Δ546 EOM model 3588 superimposed on top. SAP97 546r2 (magenta) fixed, α-Helices of the U4 linker (cyan) and the U4 linker domain (gold).

7.2.4 SAP97 Δ546 SAXS Analysis Discussion

The modelling of SAP97 Δ546 worked well using both I-TASSER and MODELLER. For the low resolution approach the overall shape of the model was all that was required for the rigid body modelling and superimposition of the GASBOR envelopes. The I-TASSER model using two PSD95 structures as templates gave the overall best model.

BUNCH is the first starting point for rigid body modelling as it attempts to find the best solution to the scattering data. This method works well for multi domain proteins that are stable in conformation and have few flexible regions. For Δ546 BUNCH struggled with the positioning of both the Hook and the U4 linker region and the *ab initio* modelling of the regions, often wrapping the domains around other domains. One good model was produced that overlaid well with the PSD95 crystal structure.

EOM was developed to analyse flexible proteins or proteins with flexible linker regions. The models generated using EOM were mostly compact (60 %) in structure as expected from the PSD95 model but the U4 linker also showed movement and elongation.

When the different EOM models along with the BUNCH model were superimposed into the individual GASBOR envelopes different models fitted different envelopes. This suggests on this data alone that the movement and flexibility U4 linker domain is real and not an artefact of the EOM, highlighting that Δ546 is a dynamic system. This movement would also account for why there are so many differences in the GASBOR envelopes and why BUNCH was not very successful at modelling Δ546.

7.3 SAP97 Δ 461 Results

7.3.1 SAP97 Δ461 Scattering Results

As with Δ546 above the SAXS data was collected by batch method for 1, 2, and 3 mg/ml concentrations, and then initially analysed by the synchrotron principal beamline scientist, Dr Javier Pérez the data was averaged and the background subtracted. The scattering data was then further processed using the program GNOM, and D_{max} 150 Å and R_g 31.9 Å were derived for SAP97 Δ461 (Figure 7.3.1).

The P(r) distribution shows a broad peak that indicates a spherical shape followed by a hump from ~60 – 120 Å. This suggests that there is another area that is quite elongated possibly with a smaller domain present. The D_{max} is quite large for just the addition of a PDZ domain. With the P(r) almost at zero at 120 Å it is most likely that this is the actual D_{max}, the further elongation to 150 Å is just an effect of the GNOM processing.

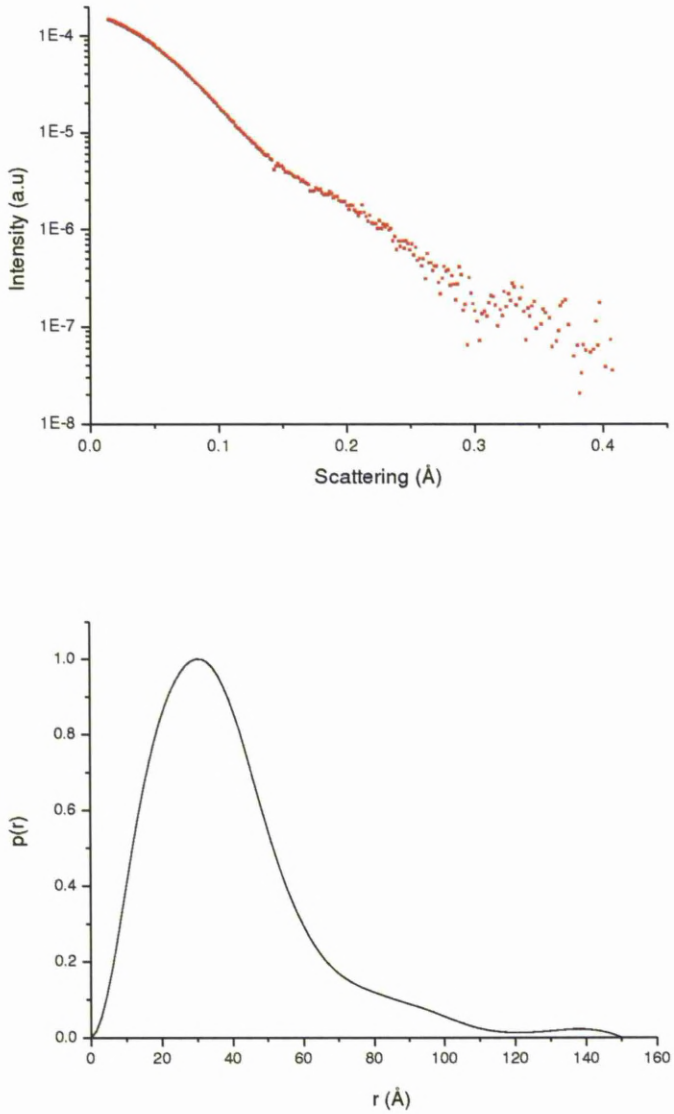


Figure 7.3.1. Small angle X-ray scattering of SAP97 Δ461. The averaged scattering data from 1, 2 and 3 mg/ml concentrations of SAP97 Δ461. The inset shows the P(r) value (regularization parameter) calculated by GNOM and gives the Rg 31.9 Å and the D_{max} 150 Å.

7.3.2 SAP97 Δ461 GASBOR and DAMAVER

Models of Δ461 were created using GASBOR (Figure 7.3.2). The variation between the individual envelopes was less than with the Δ546 envelopes. They each showed a large mass followed by a smaller mass and then a thin ribbon and a small nodule. Some of the envelopes were also noticed to be thinner than Δ 546 appearing to be all in one plane. This suggests that the PDZ3 domain would be positioned alongside the SH3-GK domain. The envelopes were then averaged (Figure 7.3.3) using the program DAMAVER (Volkov and Svergun, 2003). The averaged envelopes (n = 10 no rejections) smoothed out any differences between the first and second large masses and widened the ribbon area and increased the volume of the nodule. It is possible that the nodule volume is caused by the hexahistag tail and the other domains are within the main body of the envelope.

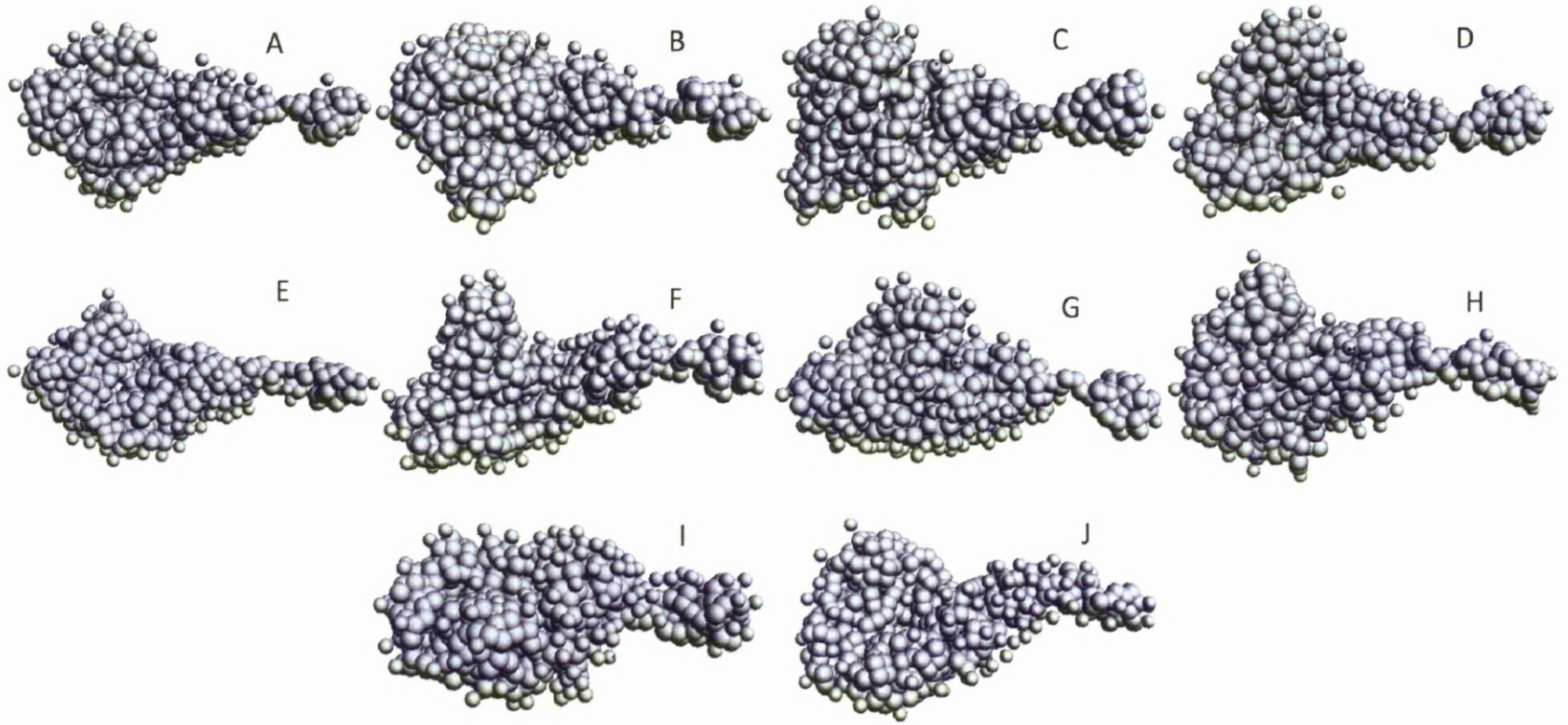


Figure 7.3.2. Individual model envelopes for SAP97 Δ461. 3D envelopes generated from 1D scattering data using program GASBOR, A-J show the 10 replicates generated from 10 individual runs. Each model is approximately 90 Å in length.

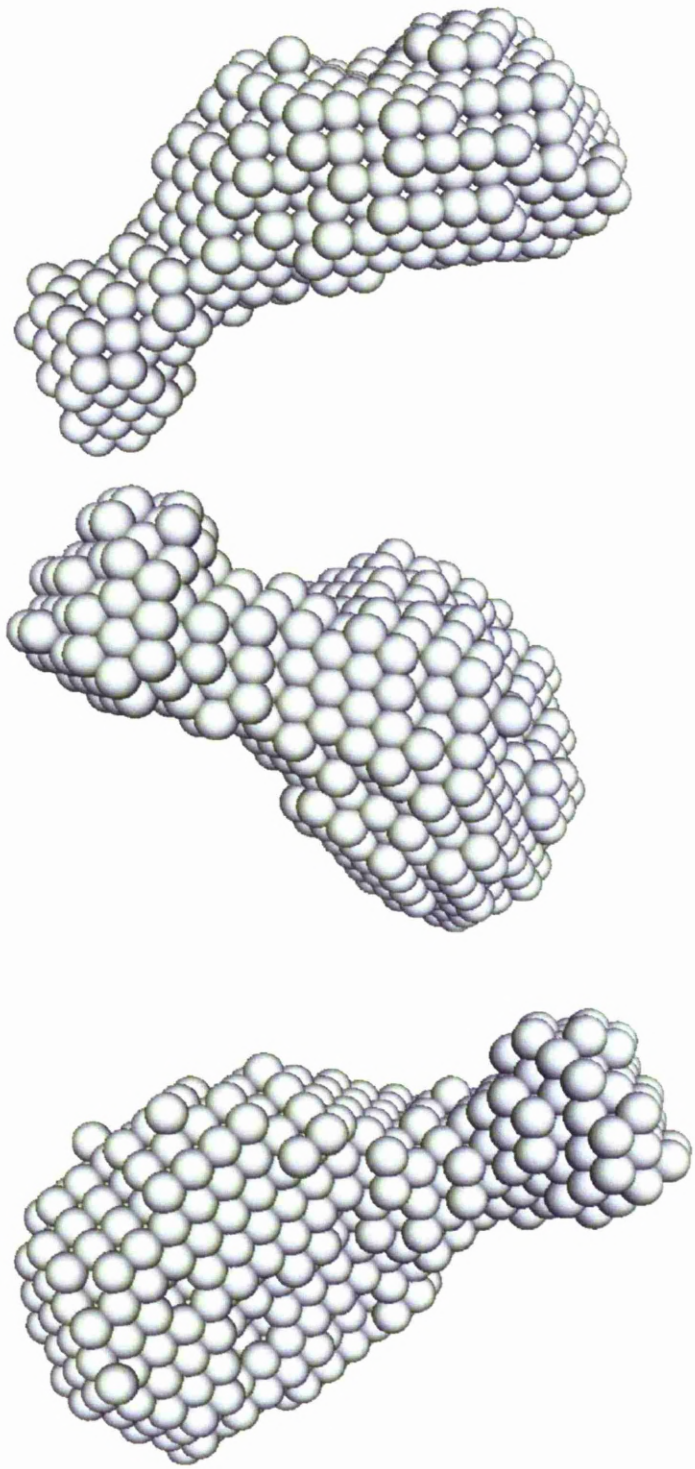


Figure 7.3.3. Averaged model envelope for SAP97 Δ461. Three orientations of the 3D envelope generated from 1D scattering data using program GASBOR and averaged using program DAMAVER (n=10).

7.3.3 SAP97 Δ461 EOM Rigid Body Modelling

The rigid body modeller, BUNCH was first tested with Δ461 however, the resultant models were not physiologically possible; the PDZ3 domain would often overlap with the SH3 domain possibly due to the flexibility in the U4 linker region seen in Δ546. Therefore, EOM modelling was used to create high resolution template structures. Initially the EOM used separate GK and SH3 domains but this did not work as EOM created many conformers in which both domains were strung apart by the Hook linker region. This led to the use of the previously defined SAP97 546r2 BUNCH model. This structure provided a good model that matched up well with the PDS95 crystal structure. Small adjustments to the model were required, the hexahistag was removed to match with the Δ461 sequence and the two helixes that contributed to the U4 linker were also removed to allow EOM to appropriately model the flexible regions. In addition, the high resolution SAP97 PDZ3 crystal structure that had previously been determined by Cabral et al 1996 was also used (1PDR).

The SAP97 Δ461 EOM frequency distribution (Figure 7.3.4) shows the pool of 10000 structures spread from 25 to 48 Å. Most strikingly however is that two distributions of Δ461 are present, a compact structure spread from 24 to 30 Å with an average R_g ~28 Å that contains 59.6 % of all conformers and a second distribution peak 39 to 48 Å with an average R_g ~45 Å that contains 38.1 % of all conformers. This shows that Δ461 exists in two forms spending over half its time in a compact form and the rest elongated. The compact form has an R_g similar to Δ546 that suggests that the PDZ3 domain is very close to the GK-SH3 domains.

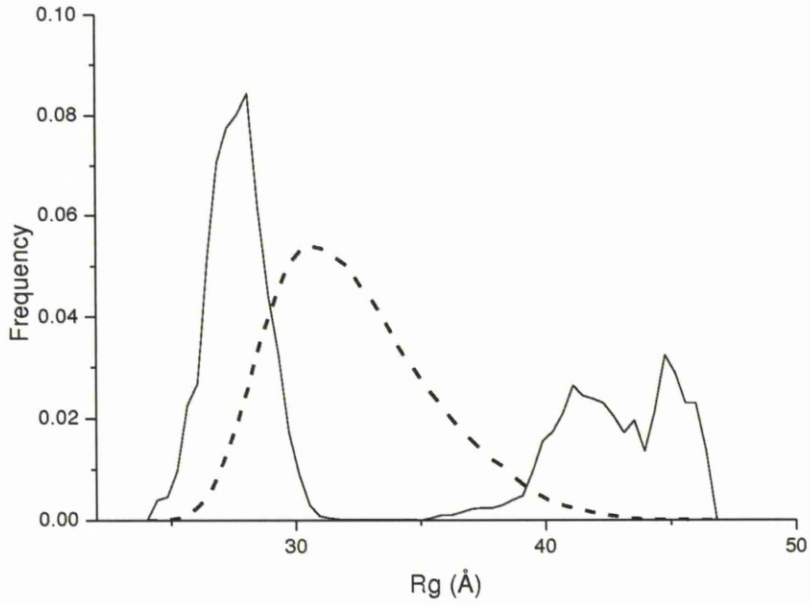


Figure 7.3.4. Analysis of SAP97 Δ461 by EOM. A frequency distribution of the SAP97 Δ461 conformers as a function of R_g. The black dashed line shows the distribution of the initial 10000 pool conformers and the black line the distribution of the analysed Δ461 ensemble.

In total, the ensemble generated 13 conformers, of these 8 were compact (Figure 7.3.5a) and 5 were elongated (Figure 7.3.5b). The compact forms showed PDZ3 restricted about the SH3 domain with the U4 linker tightly wrapped beside the PDZ3 domain. The elongated forms showed the U4 linker region fully extended with the PDZ3 domain at the end of the linker.

When the intensities plotted as a function of momentum of transfer (Figure 7.3.6) the best fit gave a good fit for the lowest angle but overall showed a poor fit, χ^2 14.3, this was because at the increased scattering levels, between 0.3 and 0.4 Å. This indicates possible inaccuracies with background subtraction during the initial

processing of the scattering. Unfortunately, this can only be remedied at the synchrotron using the "Foxtrot" program.

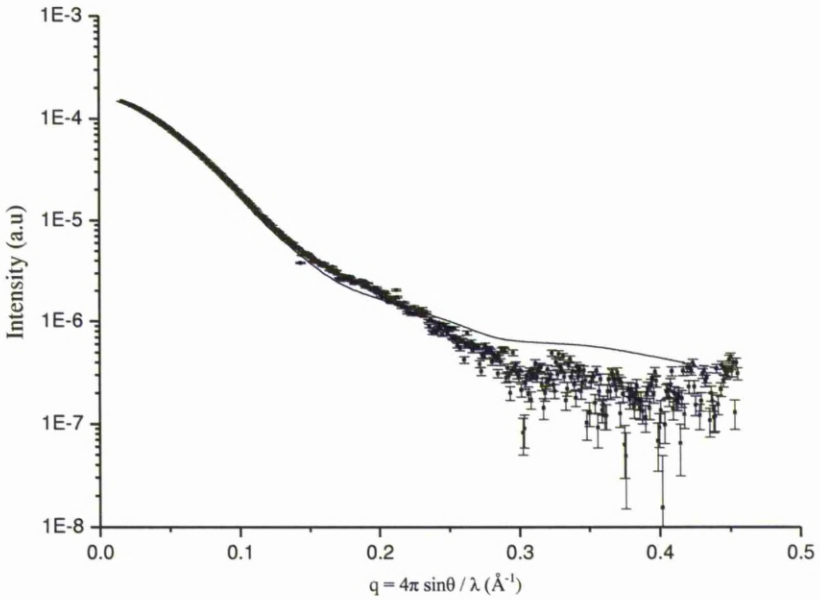


Figure 7.3.6. Analysis of SAP97 Δ461 EOM fitting. SAXS intensities plotted as a function of momentum of transfer. Dots represent experimental data; solid line is fit obtained from the ensemble of 13 conformers

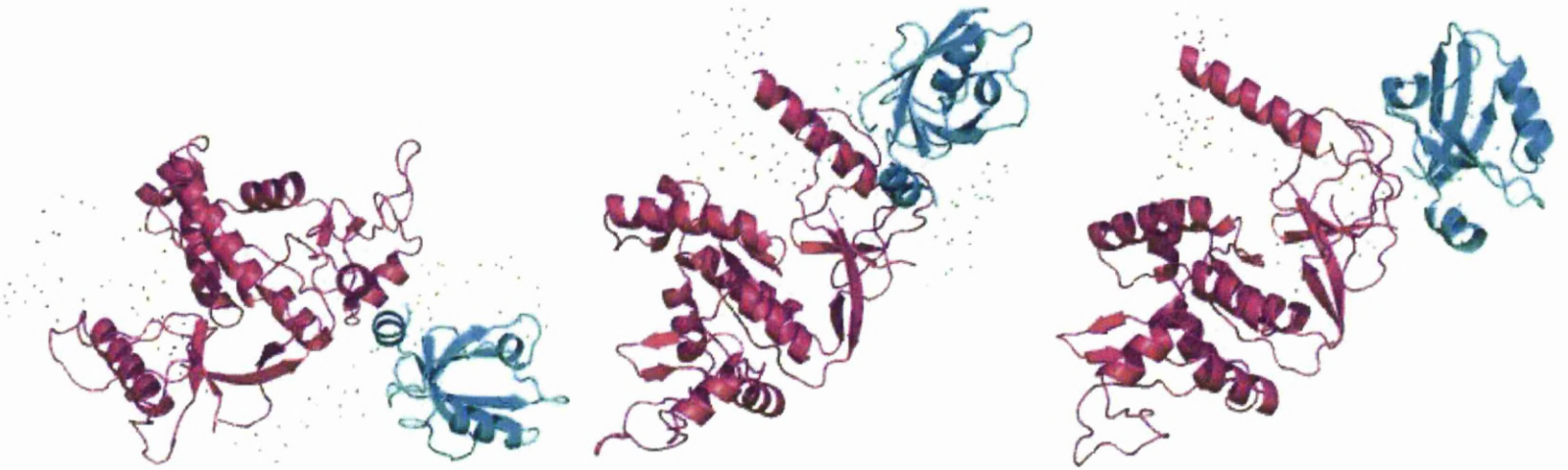


Figure 7.3.5a. SAP97 Δ461 by EOM models. A selection of the compact EOM models used SAP97 Δ546r2 BUNCH model (magenta) and PDZ3 (1PDR, cyan) in the ensemble. Each dot represents an amino acid (magenta, Hook and U4 linker green).

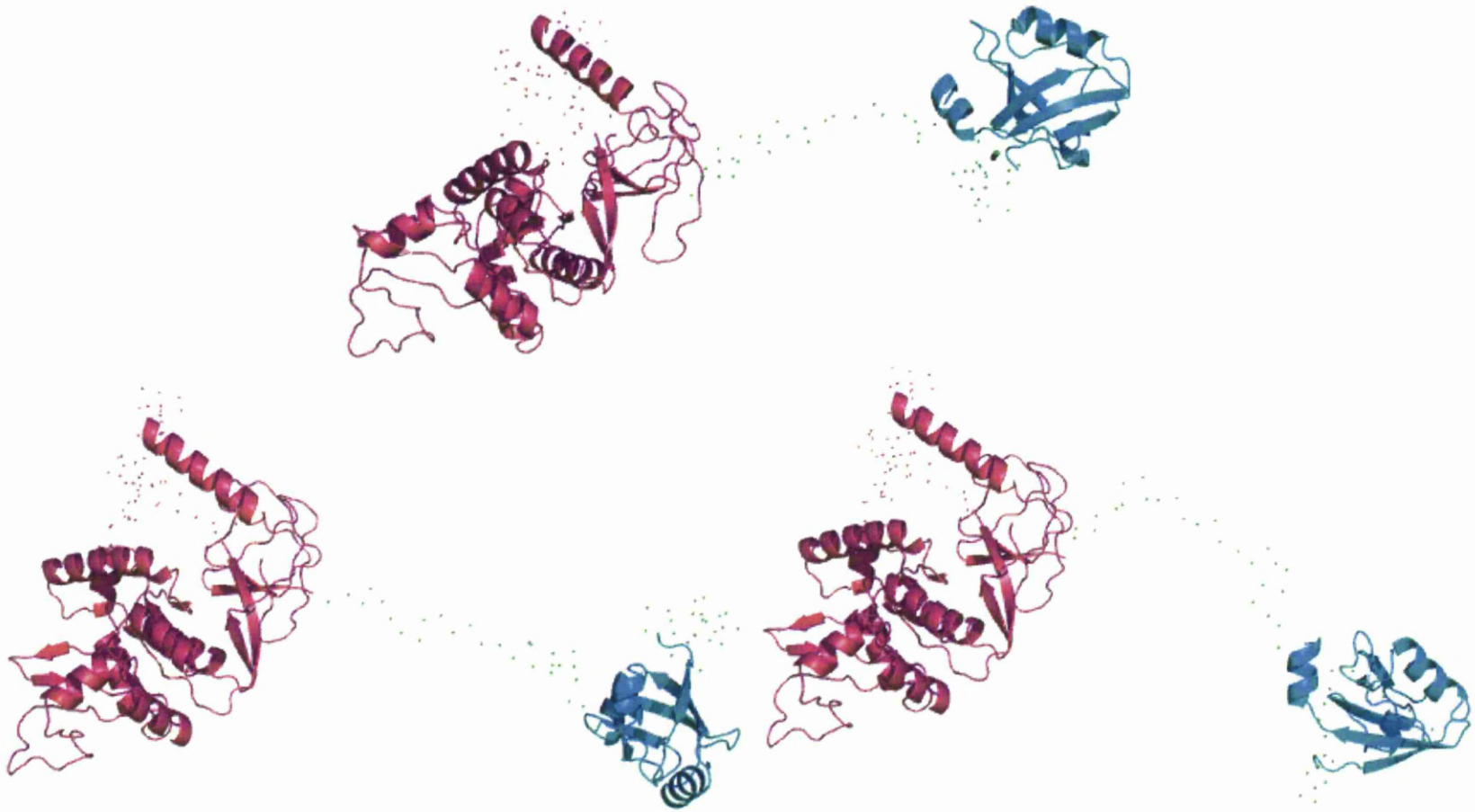


Figure 7.3.5b. SAP97 Δ461 by EOM models. A selection of the elongated EOM models used SAP97 Δ546r2 BUNCH model (magenta) and PDZ3 (1PDR, cyan) in the ensemble. Each dot represents an amino acid (magenta, Hook and U4 linker green).

7.3.5 SAP97 Δ461 EOM GASBOR Overlay

The SAP97 Δ461 EOM models using the BUNCH template were superimposed over the individual GASBOR envelopes (Figure 7.3.7) as with Δ546 earlier, the averaged GASBOR envelopes were too large for the models. Only the compact forms were overlaid as these fit the GASBOR envelopes better.

The compact EOM models found a good fit into an envelope leaving little free volume. The small nodule at the ends of the envelopes was not filled. All the overlays show that the PDZ3 needs to be tight to the SH3 domain but can move about the SH3 domain.

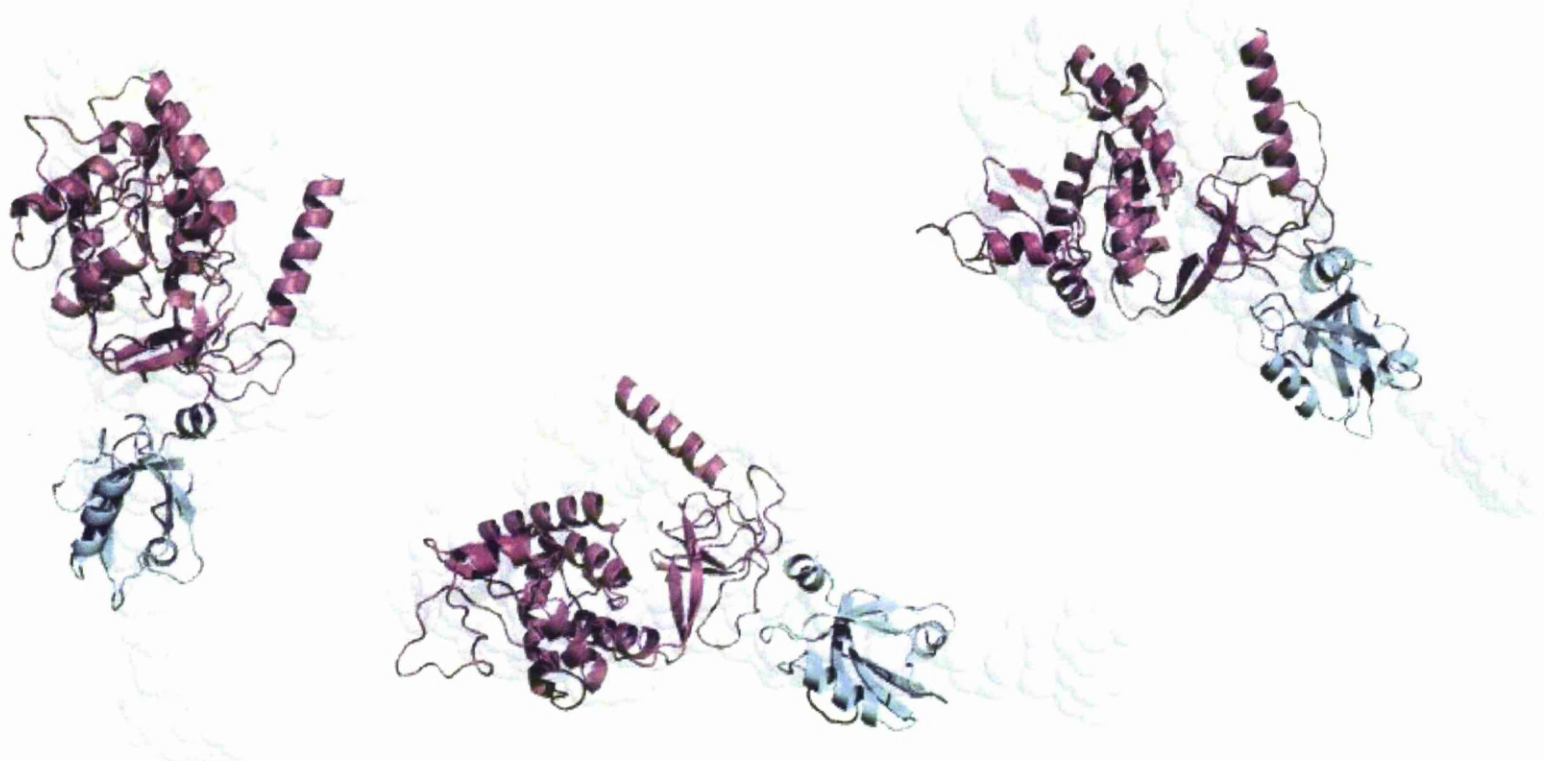


Figure 7.3.7. Shape reconstruction of SAP97 Δ461 from experimental scattering data. GASBOR envelope A is rotated with the SAP97 Δ461 EOM model 4007 superimposed on top. SAP97 Δ546r2 BUNCH (magenta) SAP97 PDZ3 (cyan).

7.3.6 SAP97 Δ461 SAXS Analysis Discussion

The structure of the three domains, PDZ3-SH3-GK together has never been solved at high or low resolution. The BUNCH modelling failed, indicating a flexible model. Using EOM together with the high resolution SAP97 PDZ3 domain and the modelled SAP97 SH3-GK domain showed that the SAP97 Δ461 can exist in two conformations a compact form and an elongated form.

The P(r) for Δ461 showed an elongated tail that gave a much larger D_{max} than expected this could be due to the elongation of the PDZ3 and U4 linker region.

The EOM also showed that the PDZ3 domain was not fixed in one position when in the compact form. It was able to move about the SH3 domain though always in one plane as shown when the models were superimposed over the flat GASBOR envelopes.

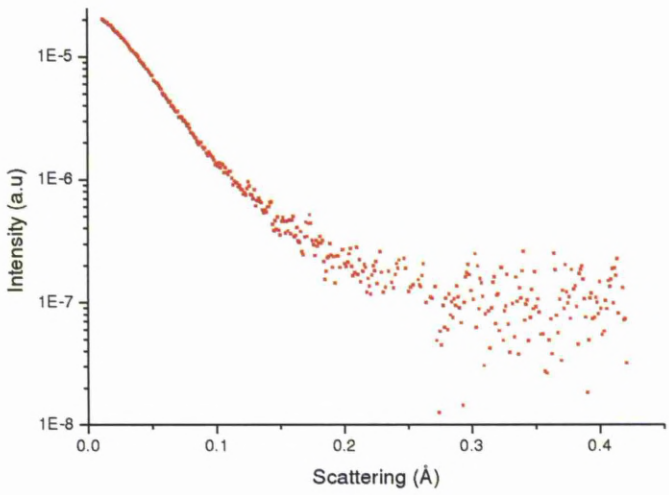
Although the EOM fitting gave a high χ^2 indicating a poor fit, possibly due to initial background subtraction, this small error in pre processing would not invalidate the two conformations that were seen.

7.4 SAP97 D70 Results

7.4.1 SAP97 Δ70 Scattering Results

Unlike Δ546 and Δ461 the data for SAP97 Δ70 was not collected through the batch method. The protein was passed through a HPLC system and eluant was continuously passed through the X-ray beam enabling the scattering data to be collected over the whole length of the column run. This allowed the scattering data of separate peaks to be collected which is useful when studying complexes or if slight degradation had occurred, which was true in this case. Thus the HPLC cleaned the sample of any slight impurities. The data from the appropriate peak was selected, averaged and the background signal subtracted as before in "Foxtrot" and using program PRIMUS. The data was then treated the same as with Δ546 and Δ461. The scattering data was then further processed using the program GNOM, and D_{max} 150 Å and R_g 42.9 Å were derived for SAP97 Δ70 (Figure 7.4.1).

The P(r) distribution shows a very broad peak with peak at about 40 Å that indicates an oval shape, followed by humps at 80 Å and 110 Å. This suggests that there is an elongated section with two areas of possible spherical volume, possibly representing the PDZ1 and 2 domains. The D_{max} is quite large, which is expected for a protein of this size.



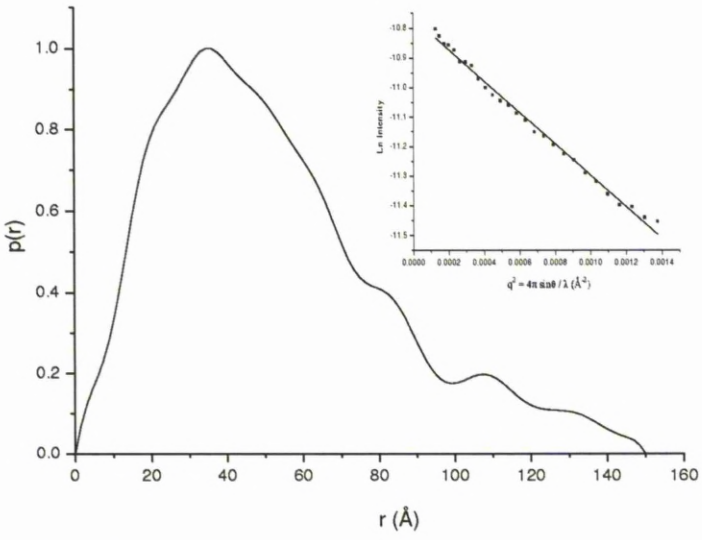


Figure 7.4.1. Small angle X-ray scattering of SAP97 Δ70. The averaged scattering data from 5 mg/ml concentration of SAP97 Δ70 passed through a HPLC. The lower plot shows the P(r) value (regularization parameter) calculated by GNOM and gives the Rg 42.9 Å and the Dmax 150 Å.

7.4.2 SAP97 Δ70 GASBOR and DAMAVER

Models of Δ70 were created using GASBOR (Figure 7.4.2). They each showed a large mass followed by a longer drawn out area that is either straight on or at 90° to the larger mass. In many cases lumps protruded from the main body of the envelope. The envelopes were then averaged (Figure 7.4.3) using the program DAMAVER (Volkov and Svergun, 2003). The averaged envelopes (n = 10 no rejections) showed the large mass, smoothed with a long arm like protrusion 90° to the larger oval. It is possible that the largest volume holds the SH3-GK domains and possibly the PDZ3 domain, the longer arm like envelope possibly holds the remaining two PDZ domains and the linker regions.

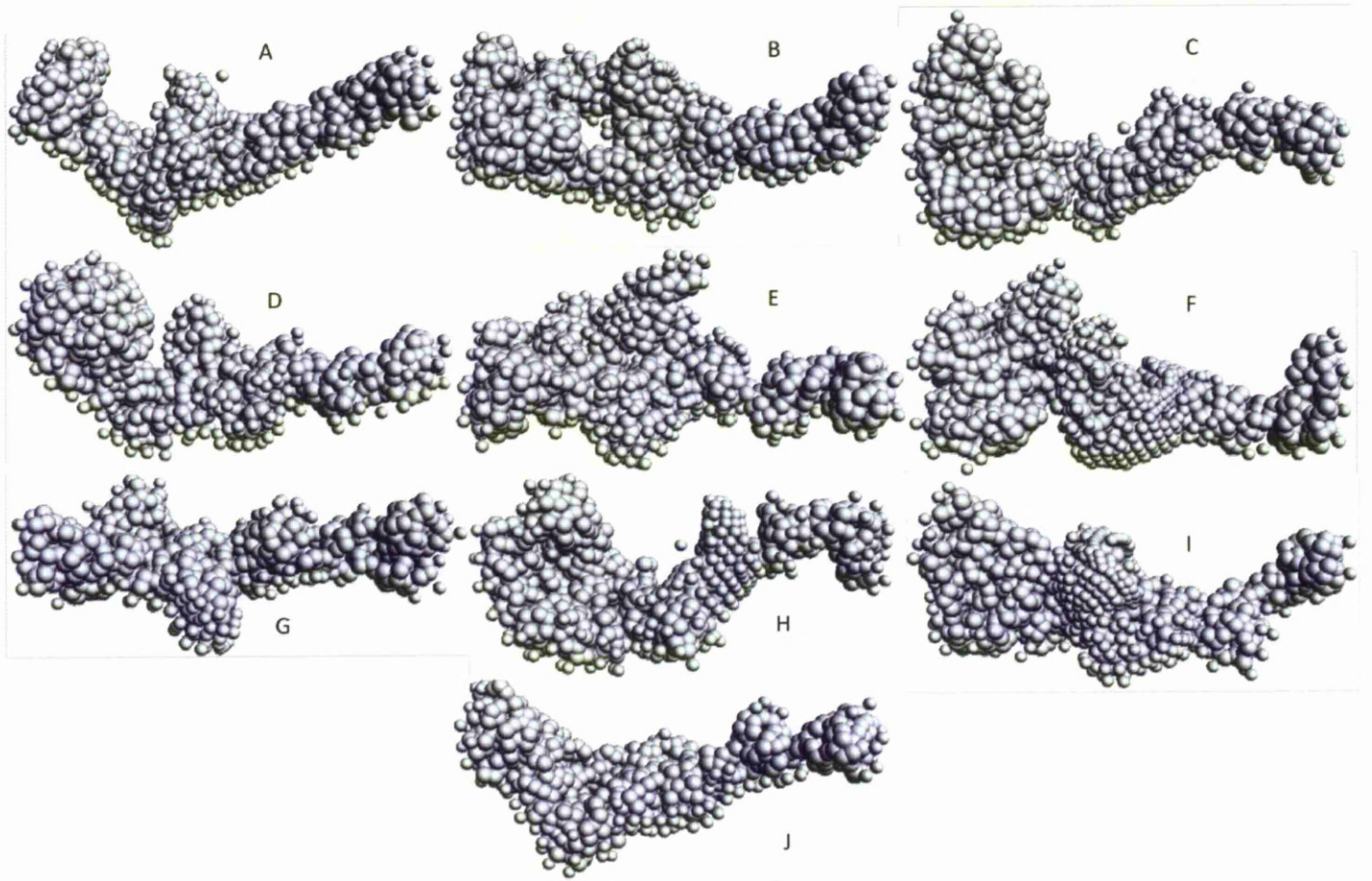


Figure 7.4.2. Individual model envelopes for SAP97 Δ70. 3D envelopes generated from 1D scattering data using program GASBOR, A-J show the 10 replicates generated from 10 individual runs. Each model is approximately 150 Å in length.

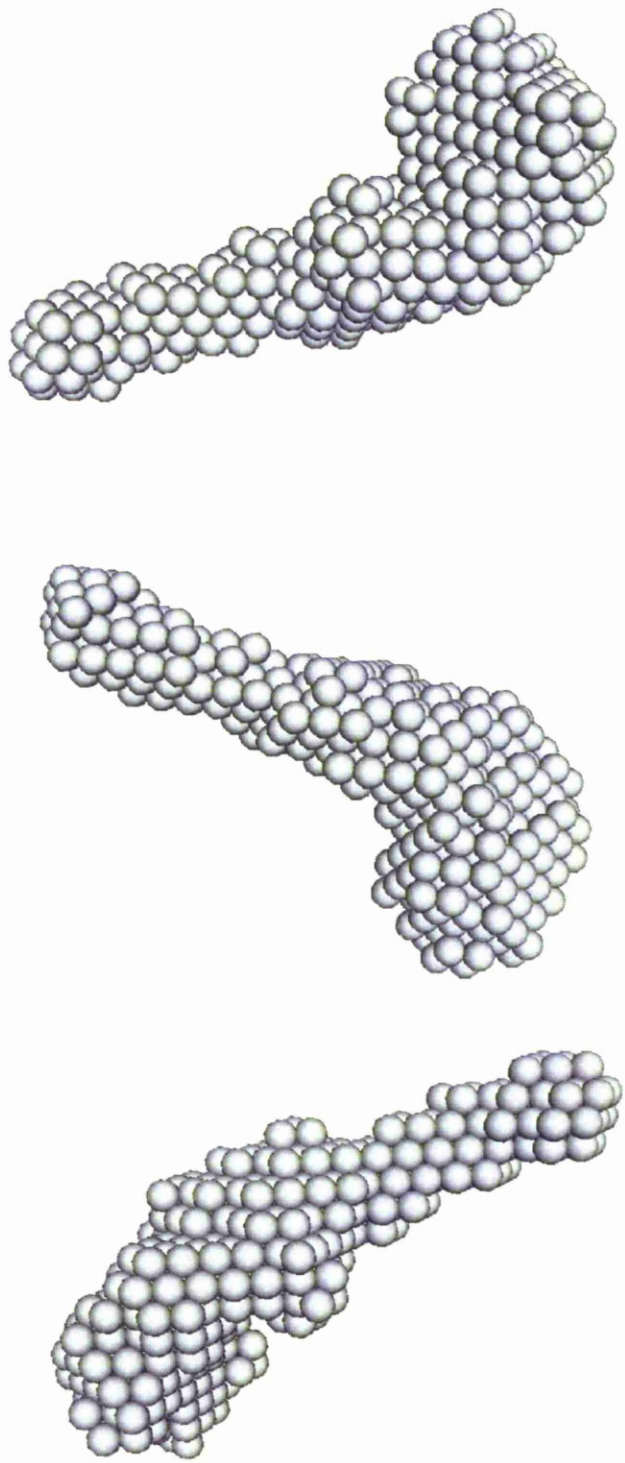


Figure 7.4.3. Averaged model envelope for SAP97 Δ70. Three orientations of the 3D envelope generated from 1D scattering data using program GASBOR and averaged using program DAMAVER (n=10).

7.4.3 SAP97 Δ70 EOM Rigid Body Modelling

As before, the EOM modelling required high resolution template structures. For consistency the previously defined SAP97 546r2 BUNCH model that was used to model Δ546 and Δ461 was used again with the SAP97 PDZ3, (1PDR, Cabral et al 1996) In addition, the recently solved high resolution crystal structure of PDZ1 and PDZ2 together (3GSL, Sainlos, Tigaret et al 2011)

The SAP97 Δ70 EOM frequency distribution (Figure 7.4.4) shows the pool of 10000 structures spread from 30 to 87 Å. Most strikingly, however, is that two distributions of Δ70 are present, a compact structure spread from 30 to 51 Å with an average R_g ~38 Å that contains 58.6 % of all conformers, a second distribution peak 60 to 87 Å with an average R_g ~73 Å that contains 35.2 % of all conformers. This shows that Δ70 exists in more than one, a compact form and a fully elongated form. There were also intermediate forms in between the compact and elongated form.

To test that this result was real, not just an artefact of the EOM program it was repeated four more times (Figure 7.4.5) and an average frequency distribution generated. The same distribution was seen for the compact form between ~30 and 51 Å.

This shows that Δ70 is in at least two states with the elongated conformation more fluid and dynamic that is probably caused by the flexibility of the linker regions which leads to more than one conformation being seen.

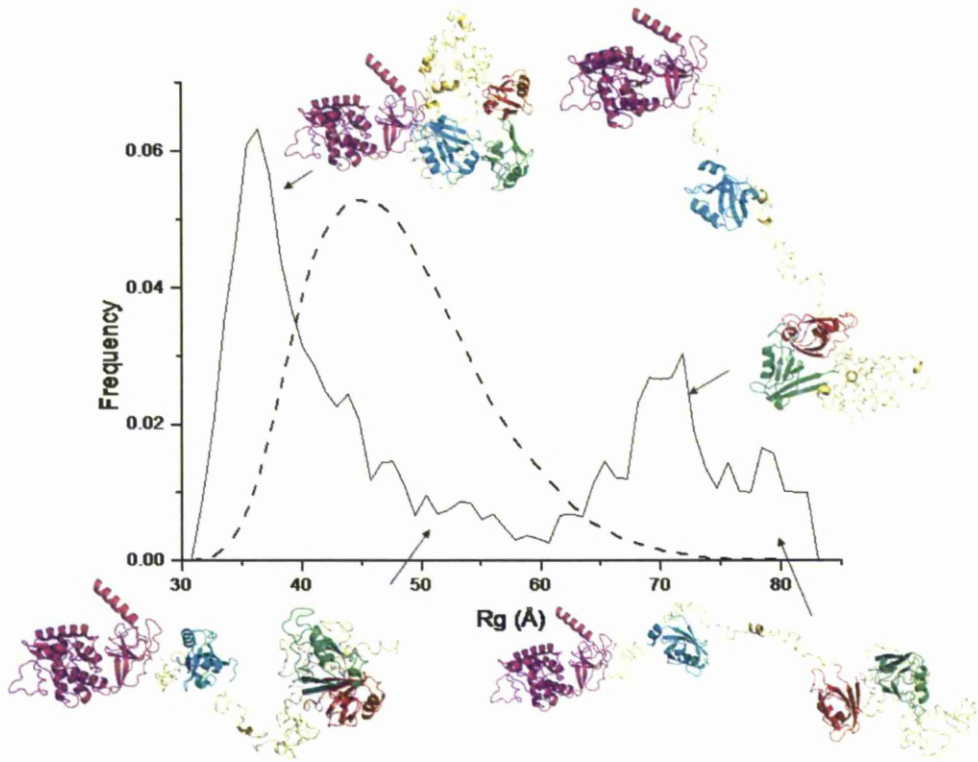


Figure 7.4.4. Analysis of SAP97 Δ70 by EOM. A frequency distribution of the SAP97 Δ70 conformers as a function of R_g. The black dashed line represents the distribution of the initial 10000 pool conformers and the solid line the distribution of the analysed Δ70 ensemble.

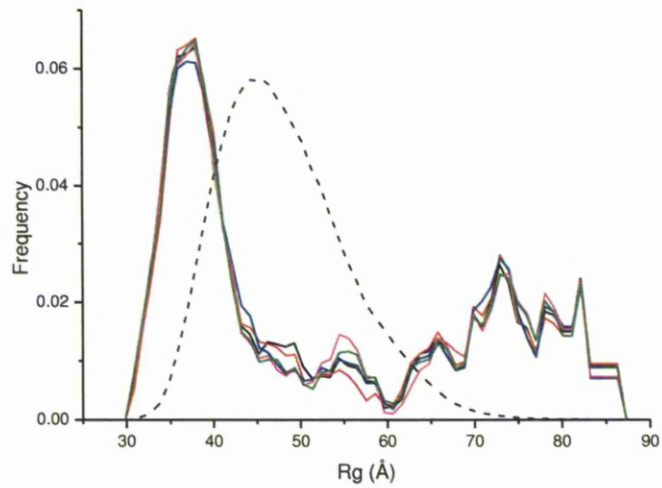


Figure 7.4.5. Analysis of SAP97 Δ70 by EOM. Frequency distributions for five SAP97 Δ70 conformers as a function of R_g. Black dash line, the distribution of the initial 10000 pool conformers and coloured lines shows the individual results for each of the five ensemble distributions.

In total, the ensemble generated 15 conformers, of these 9 were compact (Figure 7.4.6a) 5 were elongated that corresponded to the 60 to 87 Å distribution peak (Figure 7.4.6b). The compact forms showed PDZ3 domain always in close proximity to the end of the SH3 domain with the PDZ1 and 2 domains always close to each other, due to the short linker between them, but able to move about GK-SH3-PDZ3 domains. In the intermediate elongated forms the U4 linker is extended moving the PD3 domain away similar to Δ461 with the PDZ1 and 2 domains in close proximity to the PDZ3. The further extended forms show the PDZ1 and 2 domains not in proximity with the PDZ3 domain.

When the intensities plotted as a function of momentum of transfer (Figure 7.4.7) the best fitting conformer gave a good fit, χ^2 6. 9.

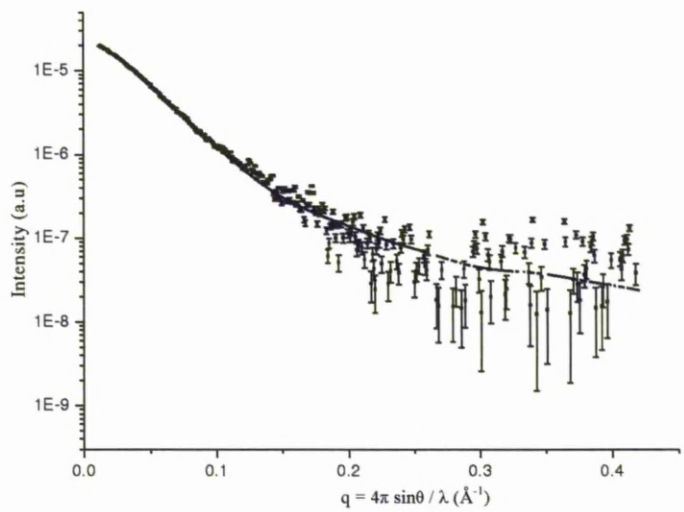


Figure 7.4.7. Analysis of SAP97 Δ70 EOM fitting. SAXS intensities plotted as a function of momentum of transfer. Dots represent experimental data; solid line is fit obtained from the ensemble of 14 conformers.

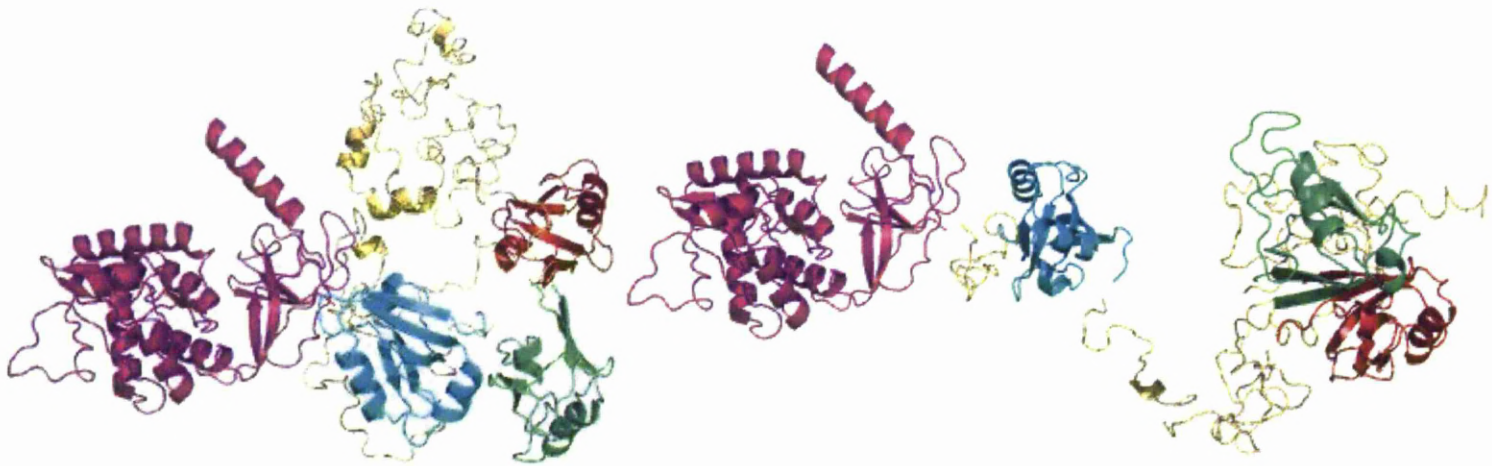


Figure 7.4.6a. SAP97 Δ70 by EOM models. Examples of the compact EOM models, left 37 Å and right 53 Å. SAP97 Δ546r2 BATCH model (magenta), PDZ3 (1PDR, cyan), PDZ2, (3GSL, red) PDZ1 (3GSL, green) and the linker regions (gold) in the ensemble.

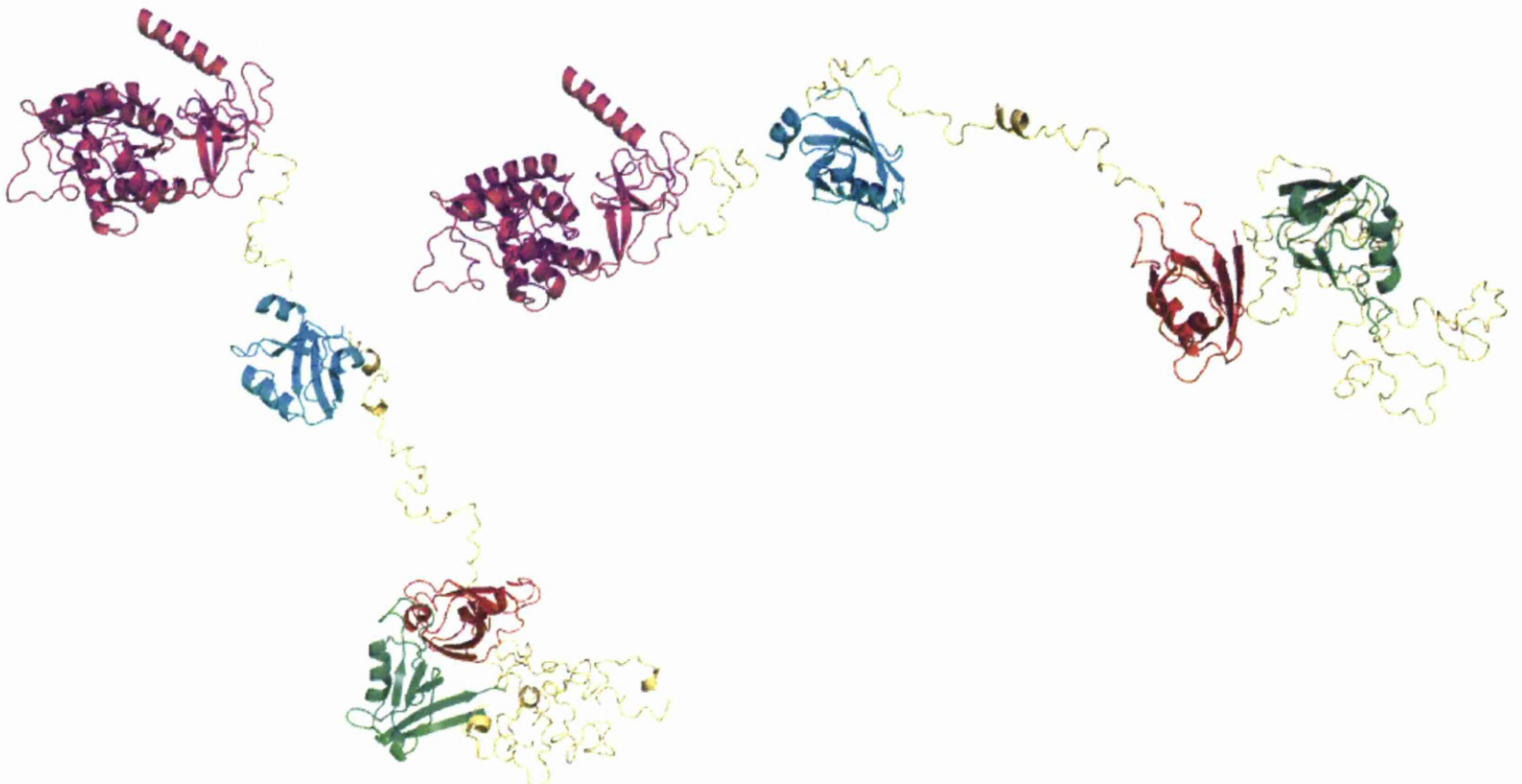


Figure 7.4.6b. SAP97 Δ70 by EOM models. The U3 and U4 fully-elongated at 70 Å (left) and semi-elongated conformer with only U3 linker extended at 80 Å (right). EOM models used SAP97 Δ546r2 BATCH model (magenta), PDZ3 (1PDR, cyan), PDZ2, (3GSL, red) PDZ1 (3GSL, green) and the linker regions (gold) in the ensemble.

7.4.4 SAP97 Δ70 EOM GASBOR Overlay

When the EOM conformers were superimposed over the GASBOR envelopes no conformer gave a good fit with any envelope. It did appear that the larger oval sphere would hold the GK– SH3 –PDZ3 domains and then the other PDZ domains may fill the arm area of the envelope but the flexibility of the linker regions caused this not to happen in this distribution.

7.4.5 SAP97 Δ70 SAXS Analysis Discussion

The addition of two extra PDZ domains and long unstructured linker regions showed an understandingly greater change to the shape of the SAP97 Δ70 than the addition of just one PDZ domain. The EOM frequency distributions showed that like with the Δ461 more than one form is possible for the protein.

Δ70 was compact with the PDZ3 domain close to the SH3 as seen in Δ461 with the PDZ1/2 domains close by or Δ70 exists in more than one conformation, most likely dynamically moving between each conformation. One of these conformations has the same compact GK-SH3-PDZ3 core but the PDZ1/2 move away extending the U3 linker between the PDZ2 and PDZ3. This linker region had been previously shown to be flexible and dynamic from previous SAXS experiments on the PDZ1/2/3 domains (Goult et al., 2007) A third possible structure sees the PDZ3 domain move away from the SH3 domain as with Δ461 but with the PDZ1/2 in close proximity as seen previously (Goult et al., 2007). A final conformation sees both the U4 and U3 linker extending to maximum creating a long drawn out

protein. I describe these different conformations as compact, semi-elongated and fully elongated (Figure 7.4.7) and examples of each were seen in the EOM frequency distributions (Figure 7.4.5b)

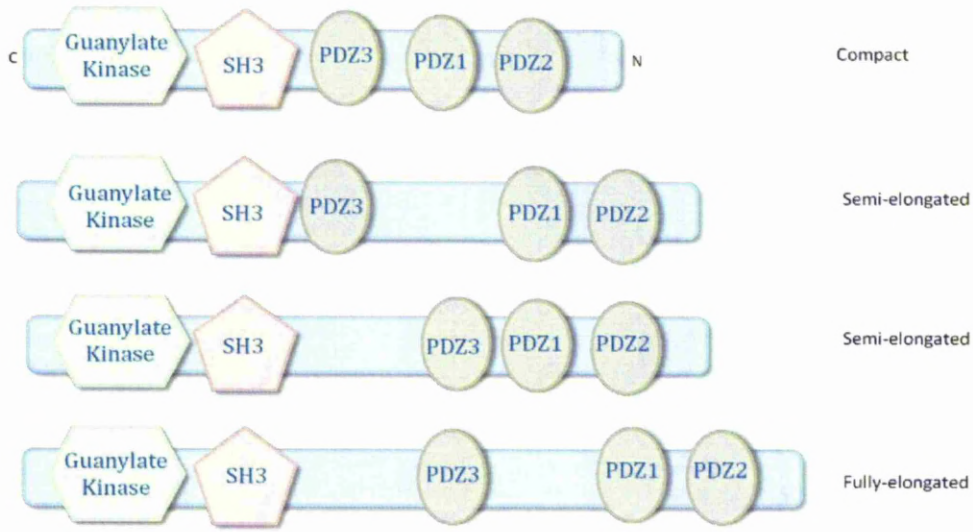


Figure 7.4.7: Schematic representation of possible SAP97 Δ70 conformations.

GASBOR was not able to produce envelopes to fit any of the EOM conformers for Δ70 that suggests the program is not suited to building envelopes for large multi domain highly flexible proteins. It might be necessary to either find another program that can handle the larger flexible proteins or new algorithms may need to be written.

7.5 SAP97 SAXS Conclusion

A range of SAXS experiments have been used to probe the conformation of the three SAP97 constructs Δ546, Δ461 and Δ70 each with varying success.

The GASBOR envelope produces a range of configurations or areas not one unique solution from the scattering data with the averaged envelope giving the most probable structure (Svergun, Petoukhov et al. 2001). Here the averaged envelopes did not fit any model but fortunately the individual envelopes were able to have individual ensemble conformers and BUNCH models superimposed upon them for Δ546 and Δ461. This was probably due to the individual GASBOR envelopes varying so much that the averaged model, created by DAMAVER was not a true representative model.

It is also probable that both GASBOR and DAMAVER are just not able to produce and average envelopes for large, highly flexible proteins.

The rigid body modelling using BUNCH gave good results for the positioning of the GK and SH3 domains for Δ546 but struggled to predict the more flexible regions often pooling many of the amino acids together. The best BUNCH model (546r2) did have the flexible regions modelled and did match well with the previously solved PSD95 crystal structure of the GK – SH3 domains.

When the rigid body modeller, EOM, a program that was designed for use with more flexible or unstructured proteins (Bernado et al., 2007) was used on Δ546 the results generated agreed in part with the BUNCH showing that the construct was compact in nature but the U4 linker was able to elongate.

With the introduction of the PDZ3 domain in Δ461 the frequency distributions created using template 546r2 GK – SH3 from BUNCH fitted the GASBOR envelopes.

The most interesting result from the EOM of Δ461 was that it existed in more than one conformation with a compact form where PDZ3 is present close to the SH3 domain and an elongated form with the PDZ3 tethered at the end of the U4 linker. This was replicated with the EOM of Δ70 where the PDZ3 domain was seen to be either next to the SH3 domain or tethered at the limit of the U4 linker. Interestingly, the U3 linker between the PDZ2 and PDZ3 domains also showed this tendency to be coiled up or strung out. This suggests that the conformations are dynamic and moving between each conformer.

This new evidence of dynamic protein domains is not unheard of. It was recently suggested that the long taught mechanism of binding by "Induced fit" may not work in every case. Instead "Conformational Selection" might be more relevant. In this hypothesis before binding the protein exists in an "ensemble of different conformations in dynamic equilibrium" in which the specific binding partner would interact with one of the conformations leading to a shift in the equilibrium of the whole protein to that conformation (Boehr and Wright, 2008). This new Conformational Selection model would fit with the modelling data generated by EOM for Δ461 and Δ70.

Another similar binding mechanism that would benefit from a protein with flexible linkers is the "Fly-casting mechanism" here the unstructured or unfolded protein has a greater capture radius or a greater reach for its binding partners than a constrained protein with restricted movement, and this allows the unstructured protein to interact weakly at first with its binding partner at long distance and then fold back bringing the binding partner closer and binding stronger (Shoemaker et al., 2000). Although this mechanism would utilize some of the dynamics of the system in SAP97 each of the PDZ domains has distinct binding partners that bind tightly in the first instance. However, if the fly-casting mechanism was adapted slightly so that each domain could be "cast-out" to enable them to locate binding partners, once bound then the protein could "reel-in" or fold back to a compact structure that would enable interactions between binding partners or help to create the scaffold. So the mechanism would be more akin to "net-fishing" or "barbed fishing" or if the three PDZ domains are taken into account then "multiple line-fishing" might be more relevant to carry on the mechanistic metaphor.

To prove if either of these mechanisms would be applicable to SAP97 specific binding experiments and further measurements would have to be carried.

In 2000, Wu et al proposed a model of SAP97 using computational molecular modelling with the available crystal structures at that time. Their data suggested that SAP97 was found in a compact structure that allowed intramolecular interactions between the N-terminal domain and the GK domain. They suggested that this compactness and / or the intramolecular interactions would interfere or block certain domain binding sites and that protein interactions might be required to

change the shape and allow further protein-protein interactions. They went on to describe a situation where it would be necessary that some of the domains would have to disassociate from the compact structure to interact with other proteins thus leading to the formation of many possible conformational states which may help SAP97 to act as a scaffold at the synaptic junctions.

The results generated here do back up some of this earlier modelling work. SAP97 Δ70 was shown to form both compact conformation and elongated conformations. However, no evidence of intramolecular interactions between the N-terminus and the GK or SH3 domain were observed. This may be caused by the clipping of the SAP97 molecule at Δ70 where Wu et al 2000 had no deletions or that different buffer conditions or binding partners are required.

The SAXS collection processing through a combination of GNOM, GASBOR, BUNCH and EOM has given much new experimentally derived information in relation to how SAP97 behaves in solution, showing that it is dynamic and can form multiple different conformations that may allow SAP97 to regulate the assembly of MAGUK complexes at the cell membrane.

CHAPTER 8

CONCLUSIONS

The work carried out in this thesis has increased the understanding of the individual proteins AKAP79 and SAP97 their relationship with each other and their interactions with other proteins such as calmodulin and Kir 2.1. The work has also thrown up more questions and some interesting results.

AKAP79 has been shown to be a difficult protein to work with. The main explanation for this is that AKAP79 is largely unstructured as shown through NMR and SECMALLS experiments. The expressed protein was hence susceptible to in-cell proteolysis. The fact that AKAP79 is unfolded may actually be necessary for its function. With no steric hindrances the protein is able to bind and possibly coil around its many binding partners or to bind many of the binding partners simultaneously.

The use of SUMO as a cleavable fusion protein enabled the purification of the AKAP79 M domain. The AKAP79 M domain had previously been shown to bind to SAP97 (Colledge et al., 2000). However, no interactions were detected through NMR and ITC. This does not dismiss the previous findings, but does indicate that the system is far more complicated than first thought and that further work is required to solve it. It is possible that intramolecular interactions within SAP97 as supported by Wu et al (2000) are required before binding or that another binding partner for either SAP97 or AKAP79 is crucial to form a tripartite complex.

AKAP79 was shown to interact with calmodulin, confirming previous studies (Faux and Scott, 1997). However, in this work, it was not possible to completely

determine the high-resolution structure of complex due to complications with its NMR spectra. It appears from the NMR data collected that one of two modes of binding are most likely; either the lobes of CaM enclose around the peptide or an X-shape dimer of CaM binding to two molecules of peptide (a 2:2 complex) occurs. SAXS experiments on CaM-peptide complexes would resolve this problem.
If the proposed CaM dimer binding is correct then it suggests that there would be two AKAP79 proteins bound in close proximity to act as scaffolds and to localise kinases and phosphatases.

Multi-domain SAP97 constructs were created to characterise and investigate any interdomain interactions. What was found was rather unexpected. The NMR spectra showed that there were little or no interactions between the PDZ3 domain and the SH3-GK domains and that there was movement in the PDZ123 domains from Δ70 as previously seen (Goult et al., 2007). The corresponding SAXS data then showed that the U3 and U4 linker regions were actually highly mobile that led to the formation of several different conformers of Δ461 and Δ70. This shows the power of the SAXS technique and how it can be used as a good complementary technique to NMR. Given the mobility detected by both NMR and SAXS, it is possible to predict that it would be difficult for intact SAP97 to be crystallised. The SAXS data also enabled differentiation between the different mobile conformers of the SAP97. However, one problem with SAXS was encountered; the analysis programs, notably GASBOR and DAMAVER were unable to manage the large flexible proteins which limited their usefulness for such studies, they also were unable to constrain

separate domains. This may lead to further research requiring new algorithms to be written to handle the data.

The further work carried out investigating CaM and Kir2.1 interactions with SAP97 confirmed the interactions demonstrated in previous studies by Paarmann et al. (2002) and Goult et al. (2007). Here the work went further with the CaM binding showing that the extra flexibility in the SH3-GK domains of Δ546 and Δ461 actually hindered the interactions compared to the more stable Δ70. However, the use of fluorescence with dansyl-CaM as the flurophore may not have been the best approach as this method gave binding affinities far smaller than previously reported that were not physiologically relevant due to the competition for CaM in the cell. A recent parallel study in the laboratory using ITC (unpublished data) gave a greater binding affinity more comparable to Paarmann et al. (2002) but it also observed a similar difference in affinity of CaM for Δ461 and Δ546 showing that fluorescence observations were real and in this case ITC is a more reliable method to measure the interactions.

Low resolution EM on SAP97 showed that two conformations were also present, an extended rod (65 %) and ring shaped (35 %) (Nakagawa, Futai et al, 2004). In SAP97 Δ70 the proportion in the extended form is 52 % showing that the removal of the N-terminal L27 domain influences the preference in conformation.

What functional significance the changing conformations of SAP97 from compact to elongated will have within the cells is unknown. However, Nakagawa, et al, (2004) postulated from that the two isoforms influenced the dynamics and recruitment and trafficking of proteins to the AMPER receptors. It is possible that Δ70 in the elongated form is more readily able to bind to target molecules and traffic them to the required sites of action such as Kir2.1 to the membrane, while the compact form would require further cofactors for binding. Possible extra cofactors could be AKAP79 or CaM or a combination of the two.

CHAPTER 9

FURTHER WORK

Characterising and investigating the interactions between several different proteins will always create many more questions that will require more study.

AKAP79 interactions were not fully investigated and will require further study; the investigation into the binding of AKAP79 M to CaM will require further NMR to probe the interaction using the recombinant peptides. However, if there are problems with line broadening then it may be necessary to form the complex using X-ray crystallography. This method has been shown to work previously with CaM binding interactions (Majava and Kursula 2009; de Diego, Kuper et al., 2010). The use of SAXS with the bound complex should be able to differentiate betwen a 2:2 ratio and a 1:1 interactions but mutations to the CaM binding sites may be necessary to help identify the exact binding mechanism. It would be interesting once the mechanism for CaM binding has been solved to investigate how PKC, that has been shown to bind to the same sequence (Klauck et al., 1996) interacts with the peptide.

Following the successful expression using the pOPIN-S, SUMO fusion vector, AKAP79 constructs of the C and N domains could be recloned. The AKAP79 N domain contains the three basic regions that interact with the plasma membrane (Dell'Acqua et al., 1998) and has been postulated to contain a second site of CaM interaction (Carr et al., 1992). These questions could be resolved using the same methodology as AKAP79 peptide:CaM binding interactions.

The successful expression of AKAP79 C after recloning into pOPIN-S would allow investigations into CaN binding interactions.

AKAP79 FL could be recloned into pOPIN-S and expressed to enable further binding studies with CaN and PtdIns(4,5)P_2 and PKC could investigated together.

In their current vectors little more work can be investigated with the large SAP97 constructs, except for probing binding interactions using SAXS. This would be useful to try and show whether the flexibility in the linkers and the different conformations formed, help in binding interactions. If, for instance when the PDZs are bound to their ligands only the compact, constrained model is formed then this would support my "multiple line fishing" analogy of the binding mechanism. However, it may be necessary to first reclone the SAP97 constructs. Throughout this work cleavage of the hexahistag has proved problematic from constructs cloned into the pLEICS vector, possibly caused by the flexibility in the constructs hindering the Tev protease.

A SAP97 full length construct needs to be constructed to finally answer the question of intermolecular interactions between the N-terminus (L27 domain) and the SH3-GK domain (Wu et al., 2000). This may pose problems though with the L27 domain known to cause dimer formation (Nakagawa et al., 2004).

Further work to investigate the relationship between SAP97 conformation and preferential binding will also need to be undertaken.

Much of the work here focused on the biophysical characterisation of the proteins. Eventually it would be necessary to take the findings investigate the system in a physiological environment. This could then help to evaluate the initial proposed model of SAP97-AKAP79-Kir2.1. This would require further binding interactions and in cell measurements to investigate the effects on channel activation using patch-clamping.

References

Babu, Y. S., C. E. Bugg, et al. (1988). "Structure of Calmodulin Refined at 2.2 a Resolution." Journal of Molecular Biology **204**(1): 191-204.

Berrow, N. S., D. Alderton, et al. (2007). "A versatile ligation-independent cloning method suitable for high-throughput expression screening applications." Nucleic Acids Research **35**(6): -.

Beique, F., M. Ali, et al. (2006). "Canadian guidelines for training in adult perioperative transesophageal echocardiography." Canadian Journal of Cardiology **22**(12): 1015-1027.

Berger, A., E. Schiltz, et al. (1989). "Guanylate Kinase from Saccharomyces-Cerevisiae - Isolation and Characterization, Crystallization and Preliminary-X-Ray Analysis, Amino-Acid Sequence and Comparison with Adenylate Kinases." European Journal of Biochemistry **184**(2): 433-443.

Bernado, P., E. Mylonas, et al. (2007). "Structural characterization of flexible proteins using small-angle X-ray scattering." Journal of the American Chemical Society **129**(17): 5656-5664.

Bichet, D., F. A. Haass, et al. (2003). "Merging functional studies with structures of inward-rectifier K+ channels." Nature Reviews Neuroscience **4**(12): 957-967.

Blaszczyk, J., Y. Li, et al. (2001). "Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes." Journal of Molecular Biology **307**(1): 247-257.

Boehr, D. D. and P. E. Wright (2008). "How do proteins interact?" Science **320**(5882): 1429-1430.

Brenman, J. E., D. S. Chao, et al. (1996). "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha 1-syntrophin mediated by PDZ domains." Cell **84**(5): 757-767.

Buraei, Z. and J. A. Yang (2010). "The beta Subunit of Voltage-Gated Ca(2+) Channels." Physiological Reviews **90**(4): 1461-1506.

Butt, T. R., S. C. Edavettal, et al. (2005). "SUMO fusion technology for difficult-to-express proteins." Protein Expression and Purification **43**(1): 1-9.

Cabral, J. H. M., C. Petosa, et al. (1996). "Crystal structure of a PDZ domain." Nature **382**(6592): 649-652.

Carr, D. W., Z. E. Hausken, et al. (1992). "Association of the type II cAMP-dependent protein kinase with a human thyroid RII-anchoring protein. Cloning and characterization of the RII- binding domain." J. Biol. Chem. **267**(19): 13376-13382.

Chattopadhyaya, R., W. E. Meador, et al. (1992). "Calmodulin Structure Refined at 1.7 Angstrom Resolution." Journal of Molecular Biology **228**(4): 1177-1192.

Cho, K.-O., C. A. Hunt, et al. (1992). "The rat brain postsynaptic density fraction contains a homolog of the drosophila discs-large tumor suppressor protein." Neuron **9**(5): 929-942.

Coghlan, V. M., B. A. Perrino, et al. (1995). "Association of protein kinase A and protein phosphatase 2B with a common anchoring protein." Science **267**(5194): 108-111.

Colledge, M., R. A. Dean, et al. (2000). "Targeting of PKA to Glutamate Receptors through a MAGUK-AKAP Complex." Neuron **27**(1): 107-119.

Craven, S. E. and D. S. Bredt (1998). "PDZ proteins organize synaptic signaling pathways." Cell **93**(4): 495-498.

Dart, C. and M. L. Leyland (2001). Targeting of an A Kinase-anchoring Protein, AKAP79, to an Inwardly Rectifying Potassium Channel, Kir2.1. **276:** 20499-20505.

David, G. and J. Perez (2009). "Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline." Journal of Applied Crystallography **42**: 892-900.

de Diego, I., J. Kuper, et al. (2010). "Molecular Basis of the Death-Associated Protein Kinase-Calcium/Calmodulin Regulator Complex." Sci. Signal. **3**(106): ra6-.

Dell'Acqua, M. L., M. C. Faux, et al. (1998). "Membrane-targeting sequences on AKAP79 bind phosphatidylinositol-4,5-bisphosphate." Embo Journal **17**(8): 2246-2260.

Dell'Acqua, M. L., K. L. Dodge, et al. (2002). "Mapping the Protein Phosphatase-2B Anchoring Site on AKAP79." Journal of Biological Chemistry **277**(50): 48796-48802.

Dobrosotskaya, I., R. K. Guy, et al. (1997). "MAGI-1, a membrane-associated guanylate kinase with a unique arrangement of protein-protein interaction domains." Journal of Biological Chemistry **272**(50): 31589-31597.

Doerks, T., P. Bork, et al. (2000). "L27, a novel heterodimerization domain in receptor targeting proteins Lin-2 and Lin-7." Trends in Biochemical Sciences **25**(7): 317-318.

Doyle, D. A., A. Lee, et al. (1996). "Crystal Structures of a Complexed and Peptide-Free Membrane Protein-Binding Domain: Molecular Basis of Peptide Recognition by PDZ." Cell **85**(7): 1067-1076.

Elshorst, B., M. Hennig, et al. (1999). "NMR solution structure of a complex of calmodulin with a binding peptide of the Ca2+ pump." Biochemistry **38**(38): 12320-12332.

Faux, M. C. and J. D. Scott (1997). "Regulation of the AKAP79-Protein Kinase C Interaction by Ca2+/Calmodulin." J. Biol. Chem. **272**(27): 17038-17044.

Feng, W., J. F. Long, et al. (2004). "The tetrameric L27 domain complex as an organization platform for supramolecular assemblies." Nature Structural & Molecular Biology **11**(5): 475-480.

Froehner, S. C. (1993). "Regulation of Ion Channel Distribution at Synapses." Annual Review of Neuroscience **16**: 347-368.

Funke, L., S. Dakoji, et al. (2005). "Membrane-associated guanylate kinases regulate adhesion and plasticity at cell junctions." Annu Rev Biochem **74**: 219-245.

Garcia, E. P., S. Mehta, et al. (1998). "SAP90 binds and clusters kainate receptors causing incomplete desensitization." Neuron **21**(4): 727-739.

Gardner, L. A., A. P. Naren, et al. (2007). "Assembly of an SAP97-AKAP79-cAMP-dependent Protein Kinase Scaffold at the Type 1 PSD-95/DLG/ZO1 Motif of the Human beta1-Adrenergic Receptor Generates a Receptosome Involved in Receptor Recycling and Networking." J. Biol. Chem. **282**(7): 5085-5099.

Garner, C. C., J. Nash, et al. (2000). "PDZ domains in synapse assembly and signalling." Trends in Cell Biology **10**(7): 274-280.

Greenfield, N. J. (2006). "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions." Nature Protocols **1**(6): 2527-2535.

Gold, M. G., B. Lygren, et al. (2006). "Molecular Basis of AKAP Specificity for PKA Regulatory Subunits." Molecular Cell **24**(3): 383-395.

Gold, M. G., F. D. Smith, et al. (2008). "AKAP18 Contains a Phosphoesterase Domain that Binds AMP." Journal of Molecular Biology **375**(5): 1329-1343.

Gomez, L. L., S. Alam, et al. (2002). "Regulation of A-kinase anchoring protein 79/150-cAMP-dependent protein kinase postsynaptic targeting by NMDA receptor activation of calcineurin and remodeling of dendritic actin." Journal of Neuroscience **22**(16): 7027-7044.

Gorski, J. A., L. L. Gomez, et al. (2005). "Association of an A-kinase-anchoring protein signaling scaffold with cadherin adhesion molecules in neurons and epithelial cells." Molecular Biology of the Cell **16**(8): 3574-3590.

Goult, B. T., J. D. Rapley, et al. (2007). "Small-angle X-ray scattering and NMR studies of the conformation of the PDZ region of SAP97 and its interactions with Kir2.1." Biochemistry **46**(49): 14117-14128.

Grabarek, Z. (2006). "Structural basis for diversity of the EF-hand calcium-binding proteins." Journal of Molecular Biology **359**(3): 509-525.

Hanada, T., L. H. Lin, et al. (2000). "GAKIN, a novel kinesin-like protein associates with the human homologue of the Drosophila discs large tumor suppressor in T lymphocytes." Journal of Biological Chemistry **275**(37): 28774-28784.

Harris, B. Z. and W. A. Lim (2001). "Mechanism and role of PDZ domains in signaling complex assembly." J Cell Sci **114**(18): 3219-3231.

Hidalgo, P. and A. Neely (2007). "Multiplicity of protein interactions and functions of the voltage-gated calcium channel beta-subunit." Cell Calcium **42**(4-5): 389-396.

He, J., M. Bellini, et al. (2006). "Proteomic Analysis of beta1-Adrenergic Receptor Interactions with PDZ Scaffold Proteins." **281**(5): 2820-2827.

Hoeflich, K. P. and M. Ikura (2002). "Calmodulin in action: Diversity in target recognition and activation mechanisms." Cell **108**(6): 739-742.

Hung, A. Y. and M. Sheng (2002). PDZ Domains: Structural Modules for Protein Complex Assembly. **277:** 5699-5702.

Ishida, A., I. Kameshita, et al. (2002). "Phosphorylation of calmodulin by Ca2+/calmodulin-dependent protein kinase IV." Archives of Biochemistry and Biophysics **407**(1): 72-82.

Kennedy, M. B. (1995). "Origin of Pdz (Dhr, Glgf) Domains." Trends in Biochemical Sciences **20**(9): 350-350.

Kim, E. and M. Sheng (1996). "Differential K+ Channel Clustering Activity of PSD-95 and SAP97, Two Related Membrane-associated Putative Guanylate Kinases." Neuropharmacology **35**(7): 993-1000.

Kim, E. J. and M. Sheng (2004). "PDZ domain proteins of synapses." Nature Reviews Neuroscience **5**(10): 771-781.

Kinderman, F. S., C. Kim, et al. (2006). "A Dynamic Mechanism for AKAP Binding to RII Isoforms of cAMP-Dependent Protein Kinase." Molecular Cell **24**(3): 397-408.

Klauck, T. M., M. C. Faux, et al. (1996). "Coordination of three signaling enzymes by AKAP79, a mammalian scaffold protein." Science **271**(5255): 1589-1592.

Klee, C. B., T. H. Crouch, et al. (1980). "Calmodulin." Annual Review of Biochemistry **49**: 489-515.

Koch, M. H. J., P. Vachette, et al. (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution." Quarterly Reviews of Biophysics **36**(2): 147-227.

Kornau, H.-C., P. H. Seeburg, et al. (1997). "Interaction of ion channels and receptors with PDZ domain proteins." Current Opinion in Neurobiology **7**(3): 368-373.

Leonard, A. S., M. A. Davare, et al. (1998). "SAP97 is associated with the alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor GluR1 subunit." Journal of Biological Chemistry **273**(31): 19518-19524.

Leonoudakis, D., W. S. Mailliard, et al. (2001). "Inward rectifier potassium channel Kir2.2 is associated with synapse-associated protein SAP97." Journal of Cell Science **114**(5): 987-998.

Li, S. S. C. (2005). "Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction." Biochemical Journal **390**: 641-653.

Lisman, J., H. Schulman, et al. (2002). "The molecular basis of CaMKII function in synaptic and behavioural memory." Nature Reviews Neuroscience **3**(3): 175-190.

Long, J. F., H. Tochio, et al. (2003). "Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95." Journal of Molecular Biology **327**(1): 203-214.

Lue, R. A., S. M. Marfatia, et al. (1994). "Cloning and Characterization of hdlg: The Human Homologue of the Drosophila Discs Large Tumor Suppressor Binds to Protein 4.1." **91**(21): 9818-9822.

Lye, M. F., A. S. Fanning, et al. (2010). "Insights into Regulated Ligand Binding Sites from the Structure of ZO-1 Src Homology 3-Guanylate Kinase Module." Journal of Biological Chemistry **285**(18): 13907-13917.

Malakhov, M., M. Mattern, et al. (2004). "SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins." Journal of Structural and Functional Genomics **5**(1): 75-86.

Malencik, D. A., T. S. Huang, et al. (1982). "Binding of Protein-Kinase Substrates by Fluorescently Labeled Calmodulin." Biochemical and Biophysical Research Communications **108**(1): 266-272.

Majava, V. and P. Kursula (2009). "Domain Swapping and Different Oligomeric States for the Complex Between Calmodulin and the Calmodulin-Binding Domain of Calcineurin A." PLoS ONE **4**(4): e5402.

Marcello, E., F. Gardoni, et al. (2007). "Synapse-Associated Protein-97 Mediates {alpha}-Secretase ADAM10 Trafficking and Promotes Its Activity." Journal of Neuroscience **27**(7): 1682-1691.

Marsischky, G. and J. LaBaer (2004). "Many paths to many clones: A comparative look at high-throughput cloning methods." Genome Research **14**(10B): 2020-2028.

Mauceri, D., F. Gardoni, et al. (2007). "Dual role of CaMKII-dependent SAP97 phosphorylation in mediating trafficking and insertion of NMDA receptor subunit NR2A." Journal of Neurochemistry **100**(4): 1032-1046.

McGee, A. W., S. R. Dakoji, et al. (2001). "Structure of the SH3-Guanylate Kinase Module from PSD-95 Suggests a Mechanism for Regulated Assembly of MAGUK Scaffolding Proteins." Molecular Cell **8**(6): 1291-1301.

McLaughlin, M., R. Hale, et al. (2002). "The Distribution and Function of Alternatively Spliced Insertions in hDlg." J. Biol. Chem. **277**(8): 6406-6412.

Meador, W. E., A. R. Means, et al. (1993). "Modulation of Calmodulin Plasticity in Molecular Recognition on the Basis of X-Ray Structures." Science **262**(5140): 1718-1721.

Mertens, H. D. T. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." Journal of Structural Biology **172**(1): 128-141.

Muller, S. L., M. Portwich, et al. (2005). "The tight junction protein occludin and the adherens junction protein alpha-catenin share a common interaction mechanism with ZO-1." Journal of Biological Chemistry **280**(5): 3747-3756.

Munoz, V. and L. Serrano (1997). "Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms." Biopolymers **41**(5): 495-509.

Nakagawa, T., K. Futai, et al. (2004). "Quaternary Structure, Protein Dynamics, and Synaptic Function of SAP97 Controlled by L27 Domain Interactions." Neuron **44**(3): 453-467.

Newlon, M. G., M. Roy, et al. (2001). "A novel mechanism of PKA anchoring revealed by solution structures of anchoring complexes." Embo Journal **20**(7): 1651-1662.

Nishihara, K., M. Kanemori, et al. (1998). "Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in Escherichia coli." Appl. Environ. Microbiol. **64**(5): 1694-1699.

Oliveria, S. F., M. L. Dell'Acqua, et al. (2007). "AKAP79/150 anchoring of calcineurin controls neuronal L-type Ca2+ channel activity and nuclear signaling." Neuron **55**(2): 261-275.

Opatowsky, Y., C.-C. Chen, et al. (2004). "Structural Analysis of the Voltage-Dependent Calcium Channel [beta] Subunit Functional Core and Its Complex with the [alpha]1 Interaction Domain." Neuron **42**(3): 387-399.

Paarmann, I., M. F. Lye, et al. (2008). "Structural requirements for calmodulin binding to membrane-associated guanylate kinase homologs." Protein Sci **17**(11): 1946-1954.

Paarmann, I., O. Spangenberg, et al. (2002). "Formation of complexes between Ca2+center dot calmodulin and the synapse-associated protein SAP97 requires the SH3 domain-guanylate kinase domain-connecting HOOK region." Journal of Biological Chemistry **277**(43): 40832-40838.

Park, W. S., J. Han, et al. (2008). "Physiological role of inward rectifier K+ channels in vascular smooth muscle cells." Pflugers Archiv-European Journal of Physiology **457**(1): 137-147.

Pawson, T. and J. D. Scott (1997). "Signaling through scaffold, anchoring, and adaptor proteins." Science **278**(5346): 2075-2080.

Petoukhov, M. V. and D. I. Svergun (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." Biophysical Journal **89**(2): 1237-1250.

Pinotsis, N., M. Petoukhov, et al. (2006). "Evidence for a dimeric assembly of two titin/telethonin complexes induced by the telethonin C-terminus." Journal of Structural Biology **155**(2): 239-250.

Rhoads, A. R. and F. Friedberg (1997). "Sequence motifs for calmodulin recognition." Faseb Journal **11**(5): 331-340.

Robertson, H. R., E. S. Gibson, et al. (2009). "Regulation of Postsynaptic Structure and Function by an A-Kinase Anchoring Protein-Membrane-Associated Guanylate Kinase Scaffolding Complex." J. Neurosci. **29**(24): 7929-7943.

Roy, A., A. Kucukural, et al. (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nature Protocols **5**(4): 725-738.

Sali, A. and T. L. Blundell (1993). "Comparative Protein Modeling by Satisfaction of Spatial Restraints." Journal of Molecular Biology **234**(3): 779-815.

Scholtz, J. M., Q. Hong, et al. (1991). "Parameters of Helix-Coil Transition Theory for Alanine-Based Peptides of Varying Chain Lengths in Water." Biopolymers **31**(13): 1463-1470.

Schmidt, A., D. I. Utepbergenov, et al. (2004). "Occludin binds to the SH3-hinge-GuK unit of zonula occludens protein 1: potential mechanism of tight junction regulation." Cellular and Molecular Life Sciences **61**(11): 1354-1365.

Shen, L., C. R. Weber, et al. (2008). "The tight junction protein complex undergoes rapid and continuous molecular remodeling at steady state." Journal of Cell Biology **181**(4): 683-695.

Shen, M. Y. and A. Sali (2006). "Statistical potential for assessment and prediction of protein structures." Protein Science **15**(11): 2507-2524.

Shoemaker, B. A., J. J. Portman, et al. (2000). "Speeding molecular recognition by using the folding funnel: The fly-casting mechanism." Proceedings of the National Academy of Sciences of the United States of America **97**(16): 8868-+.

Svergun, D., C. Barberato, et al. (1995). "CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates." Journal of Applied Crystallography **28**: 768-773.

Svergun, D. I. (1992). "Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria." Journal of Applied Crystallography **25**: 495-503.

Svergun, D. I. and M. H. J. Koch (2002). "Advances in structure analysis using small-angle scattering in solution." Current Opinion in Structural Biology **12**(5): 654-660.

Svergun, D. I., M. V. Petoukhov, et al. (2001). "Determination of domain structure of proteins from X-ray solution scattering." Biophysical Journal **80**(6): 2946-2953.

Svergun, D. I. (2010). "Small-angle X-ray and neutron scattering as a tool for structural systems biology." Biological Chemistry **391**(7): 737-743.

Swulius, M. T. and M. N. Waxham (2008). "Ca2+/calmodulin-dependent protein kinases." Cellular and Molecular Life Sciences **65**(17): 2637-2657.

Tavalin, S. J. (2008). "AKAP79 Selectively Enhances Protein Kinase C Regulation of GluR1 at a Ca2+-Calmodulin-dependent Protein Kinase II/Protein Kinase C Site." J. Biol. Chem. **283**(17): 11445-11452.

Tavares, G. A., E. H. Panepucci, et al. (2001). "Structural characterization of the intramolecular interaction between the SH3 and guanylate kinase domains of PSD-95." Molecular Cell **8**(6): 1313-1325.

Uhler, M. D., D. F. Carmichael, et al. (1986). "ISOLATION OF CDNA CLONES CODING FOR THE CATALYTIC SUBUNIT OF MOUSE CAMP-DEPENDENT PROTEIN-KINASE." Proceedings of the National Academy of Sciences of the United States of America **83**(5): 1300-1304.

Vetter, S. W. and E. Leclerc (2003). "Novel aspects of calmodulin target recognition and activation." European Journal of Biochemistry **270**(3): 404-414.

Vogel, H. J., J. L. Gifford, et al. (2011). "Fast methionine-based solution structure determination of calcium-calmodulin complexes." Journal of Biomolecular Nmr **50**(1): 71-81.

Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." Journal of Applied Crystallography **36**: 860-864.

von Ossowski, I., E. Oksanen, et al. (2006). "Crystal structure of the second PDZ domain of SAP97 in complex with a GluR-A C-terminal peptide." Febs Journal **273**(22): 5219-5229.

Vorherr, T., P. James, et al. (1990). "Interaction of Calmodulin with the Calmodulin Binding Domain of the Plasma-Membrane Ca-2+ Pump." Biochemistry **29**(2): 355-365.

Wang, L., A. Piserchio, et al. (2005). "Structural Characterization of the Intermolecular Interactions of Synapse-associated Protein-97 with the NR2B Subunit of N-Methyl-D-aspartate Receptors." J. Biol. Chem. **280**(29): 26992-26996.

Welling, P. A. (2008). Scaffolding Proteins in Transport Regulation. Seldin and Giebisch's The Kidney (Fourth Edition). San Diego, Academic Press: 325-341.

Willott, E., M. S. Balda, et al. (1993). "The Tight Junction Protein Zo-1 Is Homologous to the Drosophila Disks-Large Tumor-Suppressor Protein of

Septate Junctions." Proceedings of the National Academy of Sciences of the United States of America **90**(16): 7834-7838.

Wilson, M. A. and A. T. Brunger (2000). "The 1.0 angstrom crystal structure of Ca2+-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity." Journal of Molecular Biology **301**(5): 1237-1256.

Wong, W. and J. D. Scott (2004). "AKAP signalling complexes: Focal points in space and time." Nature Reviews Molecular Cell Biology **5**(12): 959-970.

Wu, H., S. M. Reuver, et al. (1998). "Subcellular targeting and cytoskeletal attachment of SAP97 to the epithelial lateral membrane." J Cell Sci **111**(16): 2365-2376.

Wu, H. J., C. Reissner, et al. (2000). "Intramolecular interactions regulate SAP97 binding to GKAP." Embo Journal **19**(21): 5740-5751.

Xia, Z. G. and D. R. Storm (2005). "The role of calmodulin as a signal integrator for synaptic plasticity." Nature Reviews Neuroscience **6**(4): 267-276.

Yap, M., S. Y. Sung, et al. (2000). "Agent managed multi database system design far the stock broking domain." 11th International Workshop on Database and Expert Systems Application, Proceedings: 471-476 1164.

Zhang, M., T. Tanaka, et al. (1995). "Calcium-Induced Conformational Transition Revealed by the Solution Structure of Apo Calmodulin." Nature Structural Biology **2**(9): 758-767.

Zhang, Y., J. Dasgupta, et al. (2007). "Structures of a Human Papillomavirus (HPV) E6 Polypeptide Bound to MAGUK Proteins: Mechanisms of Targeting Tumor Suppressors by a High-Risk HPV Oncoprotein." Journal of Virology **81**(7): 3618-3626.

Appendix 1

Buffer A

Minimal Media, 2M9. pH 7.2 in 990 ml dH_2O then autoclaved.

Reagent	Mass (in g)
Na_2HPO_4	12.5
KH_2PO_4	7.5

Buffer B

Made in 10 ml dH2O and sterile filtered through 0.2 μm syringe filter.

For labelled media substitute in ^{13}C Glucose of ^{15}N NH_4Cl

Reagent	Mass (in g)
glucose	4
NH_4Cl	1
$MgSO_4.7H_2O$	0.24
$CaCl_2.2H_2O$	0.02
Thiamin HCl	0.01

Buffer A and Buffer B are mixed and the correct antibiotic added for the expression. Buffer B can be stored at -20°C and Buffer A at room temp once autoclaved

Commonly Used Buffers

His Buffer pH 8

		Reagent	Mol Wt	mass for 1 L
1	M	NaCl	58.44	58.44
0.05	M	Tris HCl	157.6	7.88
0.00			154.2	
2	M	DTT	2	0.31
10	%	glycerol	-	100 ml

PBS *10 pH 7.3

		Reagent	Mol Wt	mass for 1 L
1.4	M	NaCl	58.44	81.82
0.02				
7	M	KCl	74.57	2.01
		Na_2HPO_4	141.9	
0.1	M		6	14.20
0.01			136.0	
8	M	KH_2PO_4	9	2.45

His Elution Buffer pH 8

		Reagent	Mol Wt	mass for 1 L
1	M	NaCl	58.44	58.44
0.05	M	Tris HCl	157.6	7.88
1	M	Imidazole	68.08	68.080
0.00			154.2	
2	M	DTT	2	0.31
10	%	glycerol	-	100 ml

Anion Exchange buffer A pH 8

		Reagent	Mol Wt	mass for 1 L
0.02	M	KCl	74.57	1.49
0.02	M	Tris HCl	157.6	3.15
0.00			372.2	
1	M	EDTA	4	0.372
0.00				
1	M	B me	78.13	0.078

Anion Exchange buffer B pH 8

		Reagent	Mol Wt	mass for 1 L
1	M	KCl	74.57	74.57
0.02	M	Tris HCl	157.6	3.15
0.00			372.2	
1	M	EDTA	4	0.372
0.00				
1	M	B me	78.13	0.078

Appendix 2 AKAP79 construct sequencing

AKAP FL sequencing 1 - 427

MKHHHHHHPMSDYDIPTTENLYFQGAMETTI
SEIHVENKDEKRSAEGSPGAERQKEKASMLCF
KRRKKAAKALKPKAGSEAADVARKCPQEAG
ASDQPEPTRGAWASLKRLVTRRKRSESSKQQ
KPLEGEMQPAINAEDADLSKKKAKSRLKIPCI
KFPRGPKRSNHSKIIEDSDCSIKVQXEAEILDI
QTQTPLNDQATKAKSTQDLSEGISQKDGDEV
CESNVSNSITSGEKVISVELGLDNGHSAIQTGT
LILEEIETIKEKQDVQPQQASPLETSETDHQQP
VLSDVPPLPAIPDQQIVEEASNSTLESAPNGKD
YESTEIVAEETKPKDTELSQESDFKENGITEEK
SKSEESKRMEPIAIIITDTEISEFDVTKSKNVPK
QFLISAENEQVGVFANDNGFEDRTSEQYETLL
IETASSLVKNAIQLSIEQLVNEMASDDNKINN
LLQ

Sequence histag and Tev cleavage site shown in Red

CLUSTAL 2.0.12 multiple sequence alignment

100.0% identity in 274 residues overlap; Score: 1385.0; Gap frequency: 0.0%

GenBank: M90359.1

```
AKAP79 FL,      1 METTISEIHVENKDEKRSAEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVAR
AKAP1FWD,       1 METTISEIHVENKDEKRSAEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVAR
                  ************************************************************

AKAP79 FL,      1 KCPQEAGASDQPEPTRGAWASLKRLVTRRKRSESSKQQKPLEGEMQPAINAEDADLSKKK
AKAP1FWD,      61 KCPQEAGASDQPEPTRGAWASLKRLVTRRKRSESSKQQKPLEGEMQPAINAEDADLSKKK
                  ************************************************************

AKAP79 FL,    121 AKSRLKIPCIKFPRGPKRSNHSKIIEDSDCSIKVQEEAEILDIQTQTPLNDQATKAKSTQ
AKAP1FWD,     121 AKSRLKIPCIKFPRGPKRSNHSKIIEDSDCSIKVQEEAEILDIQTQTPLNDQATKAKSTQ
                  ************************************************************

AKAP79 FL,    181 DLSEGISQKDGDEVCESNVSNSITSGEKVISVELGLDNGHSAIQTGTLILEEIETIKEKQ
AKAP1FWD,     181 DLSEGISQKDGDEVCESNVSNSITSGEKVISVELGLDNGHSAIQTGTLILEEIETIKEKQ
                  ************************************************************

AKAP79 FL,    241 DVQPQQASPLETSETDHQQPVLSDVPPLPAIPDQ
AKAP1FWD,     241 DVQPQQASPLETSETDHQQPVLSDVPPLPAIPDQ
                  *********************************
```

99.7% identity in 288 residues overlap; Score: 1423.0; Gap frequency: 0.0%

```
AKAP79 FL,    140 NHSKIIEDSDCSIKVQEEAEILDIQTQTPLNDQATKAKSTQDLSEGISQKDGDEVCESNV
AKAP1rev,       1 NHSKIIEDSDCSIKVQXEAEILDIQTQTPLNDQATKAKSTQDLSEGISQKDGDEVCESNV
                  *************** ******************************************

AKAP79 FL,    200 SNSITSGEKVISVELGLDNGHSAIQTGTLILEEIETIKEKQDVQPQQASPLETSETDHQQ
AKAP1rev,      61 SNSITSGEKVISVELGLDNGHSAIQTGTLILEEIETIKEKQDVQPQQASPLETSETDHQQ
                  ************************************************************

AKAP79 FL,    260 PVLSDVPPLPAIPDQQIVEEASNSTLESAPNGKDYESTEIVAEETKPKDTELSQESDFKE
AKAP1rev,     121 PVLSDVPPLPAIPDQQIVEEASNSTLESAPNGKDYESTEIVAEETKPKDTELSQESDFKE
                  ************************************************************

AKAP79 FL,    320 NGITEEKSKSEESKRMEPIAIIITDTEISEFDVTKSKNVPKQFLISAENEQVGVFANDNG
AKAP1rev,     181 NGITEEKSKSEESKRMEPIAIIITDTEISEFDVTKSKNVPKQFLISAENEQVGVFANDNG
                  ************************************************************

AKAP79 FL,    380 FEDRTSEQYETLLIETASSLVKNAIQLSIEQLVNEMASDDNKINNLLQ
AKAP1rev,     241 FEDRTSEQYETLLIETASSLVKNAIQLSIEQLVNEMASDDNKINNLLQ
                  *********************************************
```

AKAP M SUMO (152-321)

MGSSHHHHHH GSDSEVNQEA KPEVKPEVKP ETHINLKVSD GSSEIFFKIK KTTPLRRLME AFAKRQGKEM DSLRFLYDGI RIQADQTPED LDMEDNDIIE AHREQIGG IKVQEEAEILD IQTQTPLNDQ ATKAKSTQDL SEGISQKDGD EVCESNVSNS ITSGEKVISV ELGLDNGHSA IQTGTLILEE IETIKEKQDV QPQQASPLET SETDHQQPVL SDVPPLPAIP DQQIVEEASN STLESAPNGK DYESTEIVAE ETKPKDTELS QESDFKENG

Sequence histag, SUMO fusion and SUMO protease cleavage site shown in green

Theoretical pI/Mw: 4.52 / 30784.70

Cleaved

I K V Q E E A E I L D IQTQTPLNDQ ATKAKSTQDL SEGISQKDGD EVCESNVSNS ITSGEKVISV ELGLDNGHSA IQTGTLILEE IETIKEKQDV QPQQASPLET SETDHQQPVL SDVPPLPAIP DQQIVEEASN STLESAPNGK DYESTEIVAE ETKPKDTELS QESDFKENG

```
CLUSTAL 2.0.12 multiple sequence alignment


AKAP79M Template MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
AKAP79M Sequence MGSSHXHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
                 ***** **************************************************

Template         AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGIKVQEEAEILDI 120
Sequence         AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGIKVQEEAEILDI 120
                 ************************************************************

Template         QTQTPLNDQATKAKSTQDLSEGISQKDGDEVCESNVSNSITSGEKVISVELGLDNGHSAI 180
Sequence         QTQTPLNDQATKAKSTQDLSEGISQKDGDEVCESNVSNSITSGEKVISVELGLDNGHSAI 180
                 ************************************************************

Template         QTGTLILEEIETIKEKQDVQPQQASPLETSETDHQQPVLSDVPPLPAIPDQQIVEEASNS 240
Sequence         QTGTLILEEIETIKEKQDVQPQQASPLETSETDHQQPVLSDVPPLPAIPDQQIVEEASNS 240
                 ************************************************************

Template         TLESAPNGKDYESTEIVAEETKPKDTELSQESDFKENG 278
Sequence         TLESAPNGKDYESTEIVAEETKPKDTELSQESDFKENG 278
```

AKAP N pETM-11 (1-167)

M K H H H H H H P M S D Y D I P T T E N L Y F Q G A M E T T I S E I H V E
N K D E K R S A E G S P G A E R Q K E K A S M L C F K R R K K A A K A L
K P K A G S E A A D V A R K C P Q E A G A S D Q P E P T R G A W A S L K
R L V T R R K R S E S S K Q Q K P L E G E M Q P A I N A E D A D L S K K K
A K S R L K I P C I K F P R G P K R S N H S K I I E D S D C S I K V Q E E A E
I L D I Q T Q T

Sequence histag and Tev cleavage site shown in Red

```
CLUSTAL 2.0.12 multiple sequence alignment

100.0% identity in 167 residues overlap; Score: 846.0; Gap frequency: 0.0%

AKAP2,          1 METTISEIHVENKDEKRSAEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVAR
AKAP2fwd,       1 METTISEIHVENKDEKRSAEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVAR
                  ************************************************************

AKAP2,         61 KCPQEAGASDQPEPTRGAWASLKRLVTRRKRSESSKQQKPLEGEMQPAINAEDADLSKKK
AKAP2fwd,      61 KCPQEAGASDQPEPTRGAWASLKRLVTRRKRSESSKQQKPLEGEMQPAINAEDADLSKKK
                  ************************************************************

AKAP2,        121 AKSRLKIPCIKFPRGPKRSNHSKIIEDSDCSIKVQEEAEILDIQTQT
AKAP2fwd,     121 AKSRLKIPCIKFPRGPKRSNHSKIIEDSDCSIKVQEEAEILDIQTQT
                  *********************************************

100.0% identity in 167 residues overlap; Score: 846.0; Gap frequency: 0.0%

AKAP2,          1 METTISEIHVENKDEKRSAEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVAR
AKAP2rev,       1 METTISEIHVENKDEKRSAEGSPGAERQKEKASMLCFKRRKKAAKALKPKAGSEAADVAR
                  ************************************************************

AKAP2,         61 KCPQEAGASDQPEPTRGAWASLKRLVTRRKRSESSKQQKPLEGEMQPAINAEDADLSKKK
AKAP2rev,      61 KCPQEAGASDQPEPTRGAWASLKRLVTRRKRSESSKQQKPLEGEMQPAINAEDADLSKKK
                  ************************************************************

AKAP2,        121 AKSRLKIPCIKFPRGPKRSNHSKIIEDSDCSIKVQEEAEILDIQTQT
AKAP2rev,     121 AKSRLKIPCIKFPRGPKRSNHSKIIEDSDCSIKVQEEAEILDIQTQT
                  *********************************************
```

AKAP79 $_{18\text{-}52}$

MGSSHHHHHH GSDSEVNQEA KPEVKPEVKP ETHINLKVSD GSSEIFFKIK KTTPLRRLME AFAKRQGKEM DSLRFLYDGI RIQADQTPED LDMEDNDIIE AHREQIGG A E G S P G A E R Q K E K A S M L C F K R R K K A A K A L K P K A G

Sequence histag, SUMO fusion and SUMO protease cleavage site shown in green

```
Sample 18-52 Fwd (1)
Template        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
Sequence        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
                ************************************************************

Template        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAEGSPGAERQKE 120
Sequence        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAEGSPGAERQKE 120
                ************************************************************

Template        KASMLCFKRRKKAAKALKPKAG 142
Sequence        KASMLCFKRRKKAAKALKPKAG 142
                **********************
Sample 18-52 Rev (1)
Template        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
Sequence        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
                ************************************************************

Template        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAEGSPGAERQKE 120
Sequence        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAEGSPGAERQKE 120
                ************************************************************

Template        KASMLCFKRRKKAAKALKPKAG 142
Sequence        KASMLCFKRRKKAAKALKPKAG 142
                **********************
```

AKAP79 18-61

MGSSHHHHHH GSDSEVNQEA KPEVKPEVKP ETHINLKVSD GSSEIFFKIK KTTPLRRLME AFAKRQGKEM DSLRFLYDGI RIQADQTPED LDMEDNDIIE AHREQIGG A E G S P G A E R Q K E K A S **M** L C F K R R K K A A K A L K P K A G S E A A D V A R K

Sequence histag, SUMO fusion and SUMO protease cleavage site shown in green

```
Template        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
Sequence        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
                ************************************************************

Template        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAEGSPGAERQKE 120
Sequence        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAEGSPGAERQKE 120
                ************************************************************

Template        KASMLCFKRRKKAAKALKPKAGSEAADVARK 151
Sequence        KASMLCFKRRKKAAKALKPKAGSEAADVARK 151
                *******************************
```

AKAP79 24-61

MGSSHHHHHH GSDSEVNQEA KPEVKPEVKP ETHINLKVSD GSSEIFFKIK KTTPLRRLME AFAKRQGKEM DSLRFLYDGI RIQADQTPED LDMEDNDIIE AHREQIGG **A E R Q K E K A S M** L C F K R R K K A A K A L K P K A G S E A A D V A R K

```
Template        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
Sequence        MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
                ************************************************************

Template        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAERQKEKASMLC 120
Sequence        AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAERQKEKASMLC 120
                ************************************************************

Template        FKRRKKAAKALKPKAGSEAADVARK 145
Sequence        FKRRKKAAKALKPKAGSEAADVARK 145
                *************************
```

Appendix 3 SAP97 sequences

SAP97 Δ70

MHHHHHHSSGVDLGTENLYFQS MSKQ CEPVQPGNPW ESGSLSSAAV TSESLPGGLS PPVEKYRYQD EEVLPSERIS PQVPNEVLGP ELVHVSEKSL SEIENVHGFV SHSHISPIKA NPPPVLVNTD SLETPTYVNG TDADYEYEEI TLERGNSGLG FSIAGGTDNP HIGDDSSIFI TKIITGGAAA QDGRLRVNDC ILRVNEADVR DVTHSKAVEA LKEAGSIVRL YVKRRKPASE KIMEIKLIKG PKGLGFSIAG GVGNQHIPGD NSIYVTKIIE GGAAHKDGKL QIGDKLLAVN SVCLEEVTHE EAVTALKNTS DFVYLKVAKP TSMYINDGYA PPDITNSSSQ SVDNHVSPSS YLGQTPASPA RYSPISKAVL GDDEITREPR KVVLHRGSTG LGFNIVGGED GEGIFISFIL AGGPADLSGE LRKGDRIISV NSVDLRAASH EQAAAALKNA GQAVTIVAQY RPEEYSRFEA KIHDLREQMM NSSVSSGSGS LRTSQKRSLY VRALFDYDKT KDSGLPSQGL NFKFGDILHV INASDDEWWQ ARQVTPDGES DEVGVIPSKR RVEKKERARL KTVKFNSKTR GDKGSFNDKR KKNLFSRKFP FYKNKDQSEQ ETSDADQHVT SNASDSESSY RGQEEYVLSY EPVNQQEVNY TRPVIILGPM KDRVNDDLIS EFPDKFGSCV PHTTRPKRDY EVDGRDYHFV TSREQMEKDI QEHKFIEAGQ YNNHLYGTSV QSVRAVAEKG KHCILDVSGN AIKRLQIAQL YPISIFIKPK SMENIMEMNK RLTDEQARKT FERAVRLEQE FTEHFTAIVQ GDTLEDIYNQ VKQIIEEQSG PYIWVPAKEK L

Number of amino acids: 847

Molecular weight: 93874.9

Theoretical pI: 5.84

SAP97 Δ461

MHHHHHHSSGVDLGTENLYFQSMS TREP RKVVLHRGST GLGFNIVGGE DGEGIFISFI LAGGPADLSG ELRKGDRIIS VNSVDLRAAS HEQAAAALKN AGQAVTIVAQ YRPEEYSRFE AKIHDLREQM MNSSVSSGSG SLRTSQKRSL YVRALFDYDK TKDSGLPSQG LNFKFGDILH VINASDDEWW QARQVTPDGE SDEVGVIPSK RRVEKKERAR LKTVKFNSKT RGDKGSFNDK RKKNLFSRKF PFYKNKDQSE QETSDADQHV TSNASDSESS YRGQEEYVLS YEPVNQQEVN YTRPVIILGP MKDRVNDDLI SEFPDKFGSC VPHTTRPKRD YEVDGRDYHF VTSREQMEKD IQEHKFIEAG QYNNHLYGTS VQSVRAVAEK GKHCILDVSG NAIKRLQIAQ LYPISIFIKP KSMENIMEMN KRLTDEQARK TFERAVRLEQ EFTEHFTAIV QGDTLEDIYN QVKQIIEEQS GPYIWVPAKE KL

Number of amino acids: 490

Molecular weight: 55760.3

Theoretical pI: 6.65

SAP97 Δ546

MHHHHHHSSGVDLGTENLYFQSMS RPEE YSRFEAKIHD LREQMMNSSV SSGSGSLRTS QKRSLYVRAL FDYDKTKDSG LPSQGLNFKF GDILHVINAS DDEWWQARQV TPDGESDEVG VIPSKRRVEK KERARLKTVK FNSKTRGDKG SFNDKRKKNL FSRKFPFYKN KDQSEQETSD ADQHVTSNAS DSESSYRGQE EYVLSYEPVN QQEVNYTRPV IILGPMKDRV NDDLISEFPD KFGSCVPHTT RPKRDYEVDG RDYHFVTSRE QMEKDIQEHK FIEAGQYNNH LYGTSVQSVR AVAEKGKHCI LDVSGNAIKR LQIAQLYPIS IFIKPKSMEN IMEMNKRLTD EQARKTFERA VRLEQEFTEH FTAIVQGDTL EDIYNQVKQI IEEQSGPYIW VPAKEKL

Number of amino acids: 405

Molecular weight: 46962.3

Theoretical pI: 6.59

Appendix 4 NMR Chemical Shift differences

Chemical Shift Differences for CaM interaction with AKAP79 peptide

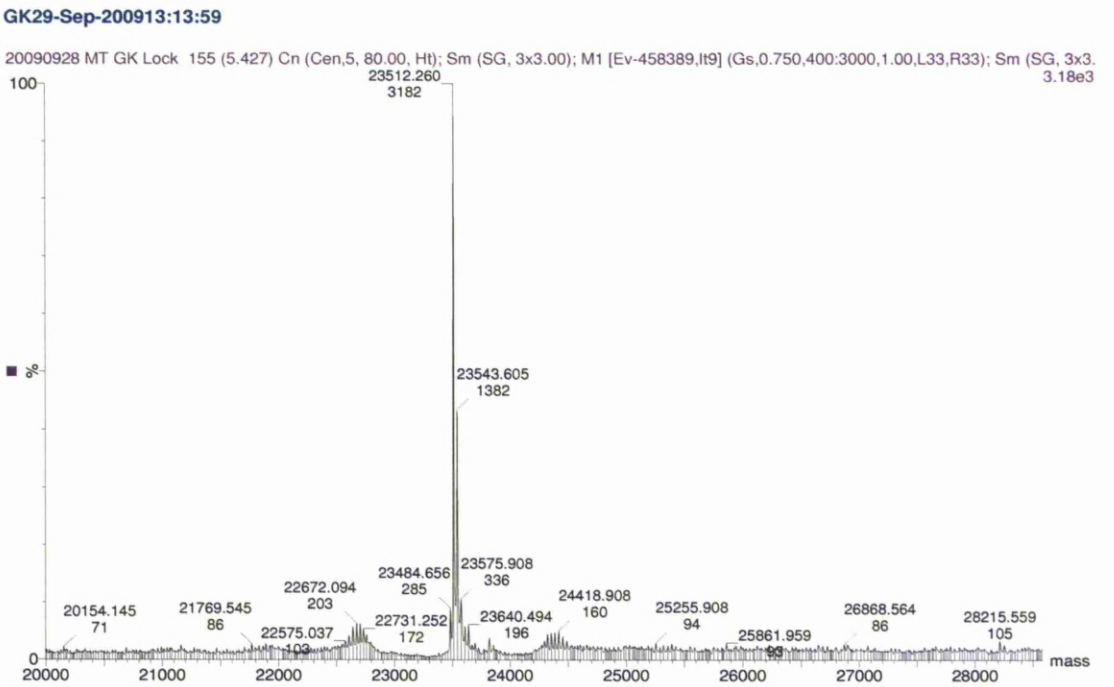
Residue	Chemical Shift	Residue	Chemical Shift	Residue	Chemical Shift	Residue	Chemical Shift
1	0	38Ser	0	75Lys	0.03829	112Leu	0
2	0	39Leu	0.01616	76Met	0	113Gly	0
3	0	40Gly	0	77Lys	0	114Glu	0
4Leu	0.02345	41Gln	0.07516	78Asp	0.06807	115Lys	0.17265
5Thr	0.07206	42Asn	0.08228	79Thr	0.08438	116Leu	0.06364
6Glu	0.02007	43	0	80Asp	0.09806	117Thr	0.04962
7Glu	0.04273	44Thr	0.06165	81Ser	0.18404	118Asp	0
8Gln	0	45Glu	0.02116	82Glu	0.01644	119Glu	0.07164
9Ile	0	46Ala	0.00867	83Glu	0.04902	120Glu	0
10Ala	0.03363	47Glu	0	84Glu	0	121Val	0.16507
11Glu	0	48Leu	0.07928	85Ile	0.06956	122Asp	0.0881
12Phe	0.05893	49Gln	0.16661	86Arg	0.16984	123Glu	0
13Lys	0.06396	50Asp	0.06248	87Glu	0	124Met	0
14Glu	0	51Met	0.05099	88Ala	0.06963	125Ile	0
15Ala	0.06358	52Ile	0.06396	89Phe	0	126Arg	0
16Phe	0.08739	53Asn	0.18431	90Arg	0.11336	127Glu	0.12096
17Ser	0.15299	54Glu	0	91Val	0.15003	128Ala	0.17821
18Leu	0.13384	55Val	0.02525	92Phe	0.04875	129Asp	0
19PheH	0	56Asp	0	93Asp	0.1339	130Ile	0.17263
20Asp	0	57Ala	0.34856	94Lys	0.14169	131Asp	0.04179
21Lys	0.00211	58Asp	0.10384	95Asp	0.06493	132Gly	0
22Asp	0	59Gly	0	96Gly	0.0209	133Asp	0.0358
23Gly	0.02302	60Asn	0	97Asn	0	134Gly	0.10089
24Asp	0.09671	61Gly	0.00361	98GlyH	0.05207	135Gln	0.0339
25Gly	0.07223	62Thr	0.09069	99Tyr	0.11586	136Val	0.01807
26Thr	0.12226	63Ile	0.16784	100Ile	0.00526	137Asn	0.11227
27Ile	0.11134	64Asp	0.08877	101Ser	0.00328	138Tyr	0
28Thr	0.05349	65Phe	0.08931	102Ala	0.0588	139Glu	0
29Thr	0.09053	66	0	103Ala	0	140Glu	0.06724
30Lys	0.05843	67Glu	0	104Glu	0	141Phe	0
31Glu	0.11066	68Phe	0	105Leu	0.1467	142Val	0.13213
32Leu	0.18341	69Leu	0	106Arg	0.17096	143Gln	0
33Gly	0.0899	70Thr	0	107	0	144Met	0
34Thr	0	71Met	0	108Val	0	145Met	0.12216
35Val	0	72Met	0.20738	109Met	0.21421	146	0
36Met	0	73Ala	0	110Thr	0.05617	147Ala	0.34112
37Arg	0	74Arg	0	111	0	148Lys	0.25846

Chemical Shift Differences for AKAP79 M and SAP97 Δ461 interaction

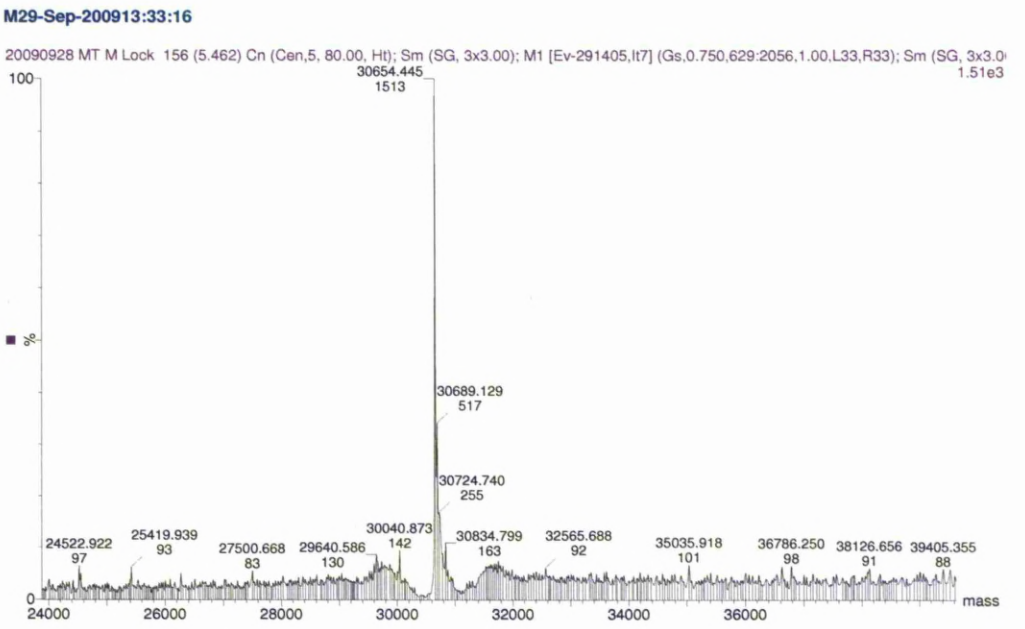
Residue	Shift Difference	Residue	Shift Difference	Residue	Shift Difference
3V	0.01425	60S	0.00429	125I	0.00883
4Q	0.01562	62E	0.00541	126V	0.01081
5E	0.00875	64G	0.00411	128E	0.01028
7A	0.01117	65L	0.00748	129A	0.00967
8E	0.00689	67N	0.00890	130S	0.00618
9I	0.00949	68G	0.01111	132S	0.00505
10L	0.00190	70S	0.00142	133T	0.00000
12I	0.00000	73Q	0.00146	134L	0.00696
13Q	0.00593	75G	0.00497	136S	0.01043
14T	0.00524	76T	0.00753	137A	0.00577
15Q	0.07360	77L	0.01081	139N	0.00801
16T	0.00147	78I	0.00000	140G	0.00618
18L	0.00918	79L	0.00898	141K	0.00000
20D	0.00000	84T	0.01039	142D	0.00000
21Q	0.00000	85I	0.00819	143Y	0.00855
22A	0.00275	91V	0.00788	144E	0.00000
23T	0.00632	94Q	0.00270	145S	0.00000
24K	0.00318	96A	0.00918	146T	0.00000
25A	0.00268	97S	0.00054	147E	0.00331
26K	0.01914	99L	0.00774	150A	0.00753
27S	0.05382	100E	0.00000	151E	0.07664
28T	0.00515	101T	0.00758	152E	0.01040
31L	0.00684	102S	0.00309	154K	0.01025
34G	0.00000	104T	0.00000	158T	0.00479
35I	0.00151	106H	0.04710	159E	0.05966
36S	0.00250	107Q	0.00000	160L	0.03091
38K	0.00618	108Q	0.02313	160L	0.00000
40G	0.01486	110V	0.02509	161S	0.06151
41D	0.00000	111L	0.00830	163E	0.00808
52I	0.00132	112S	0.00754	166F	0.01005
53T	0.00238	114V	0.01100	167K	0.01138
54S	0.00337	117L	0.00920	169N	0.00751
55G	0.00443	119A	0.00460	170G	0.00275
56E	0.01407	122D	0.00275		
59I	0.01050	123Q	0.00000		

Appendix 5 Mass spectrometry

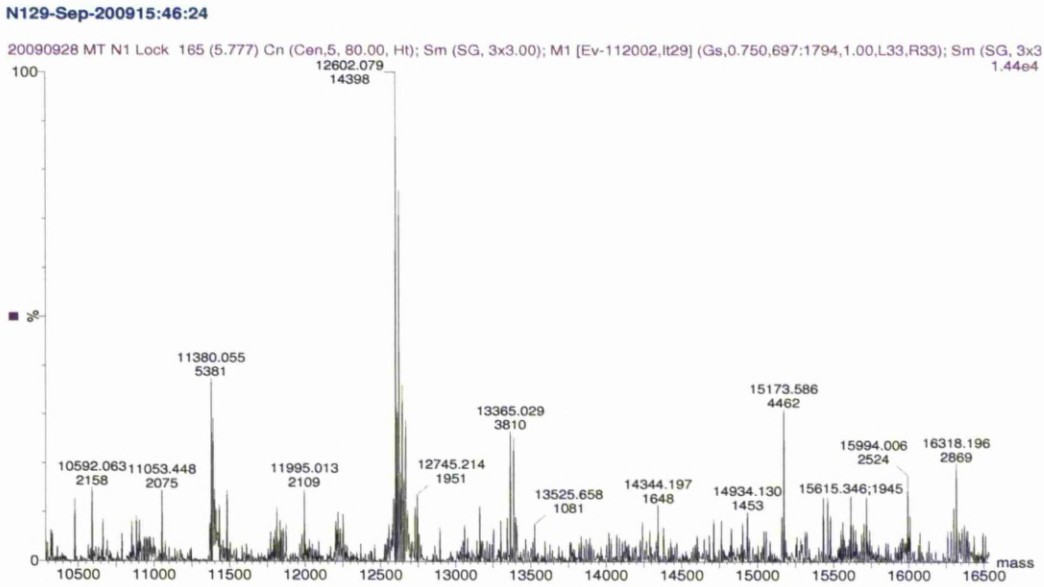
SAP97 GK



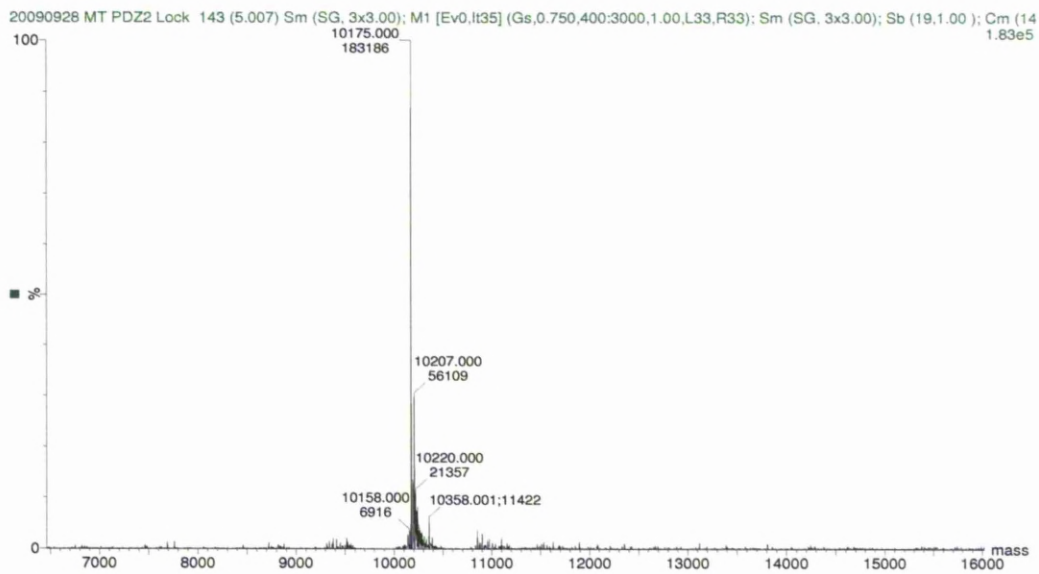
AKAP79 M



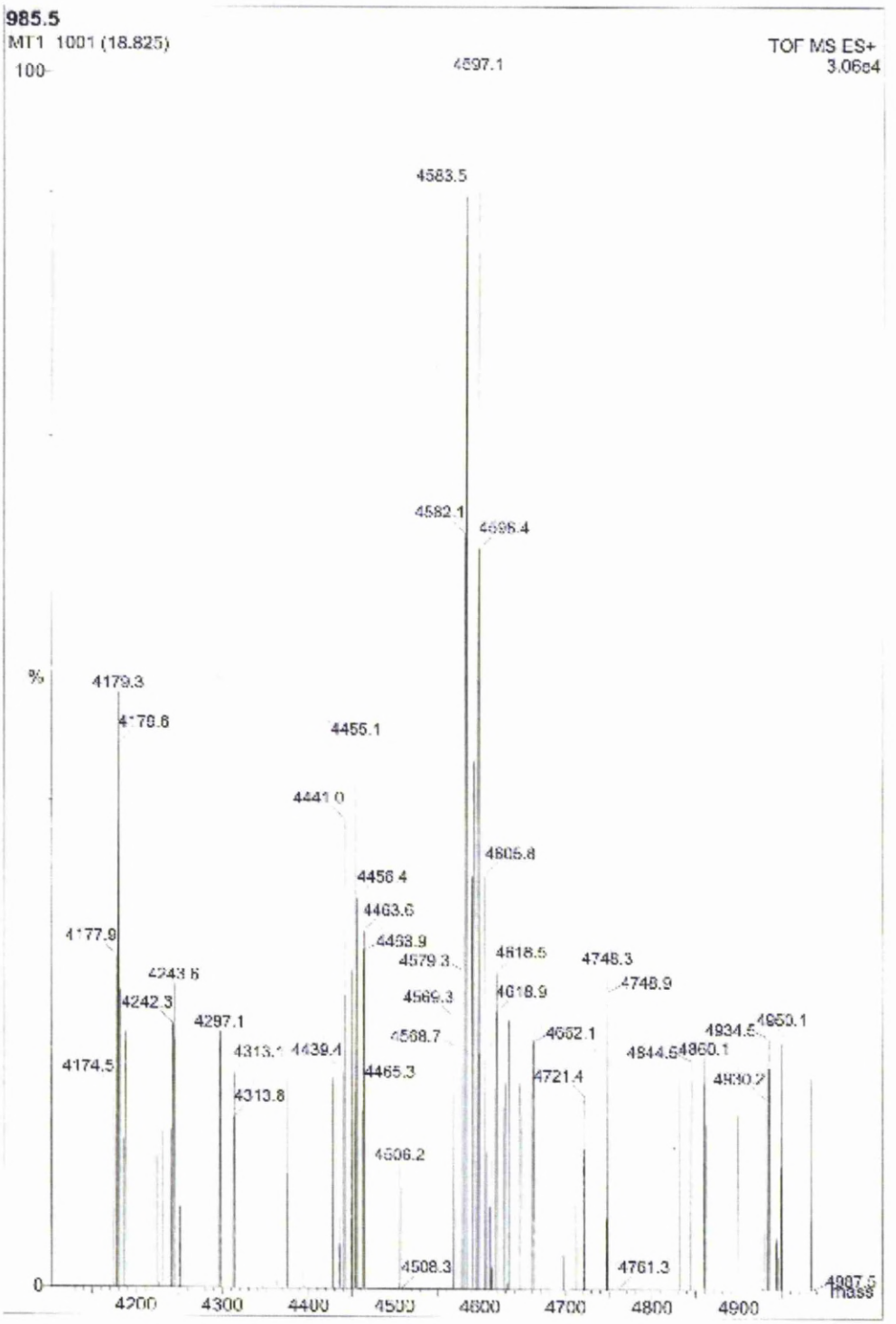
SAP97 NPDZ1



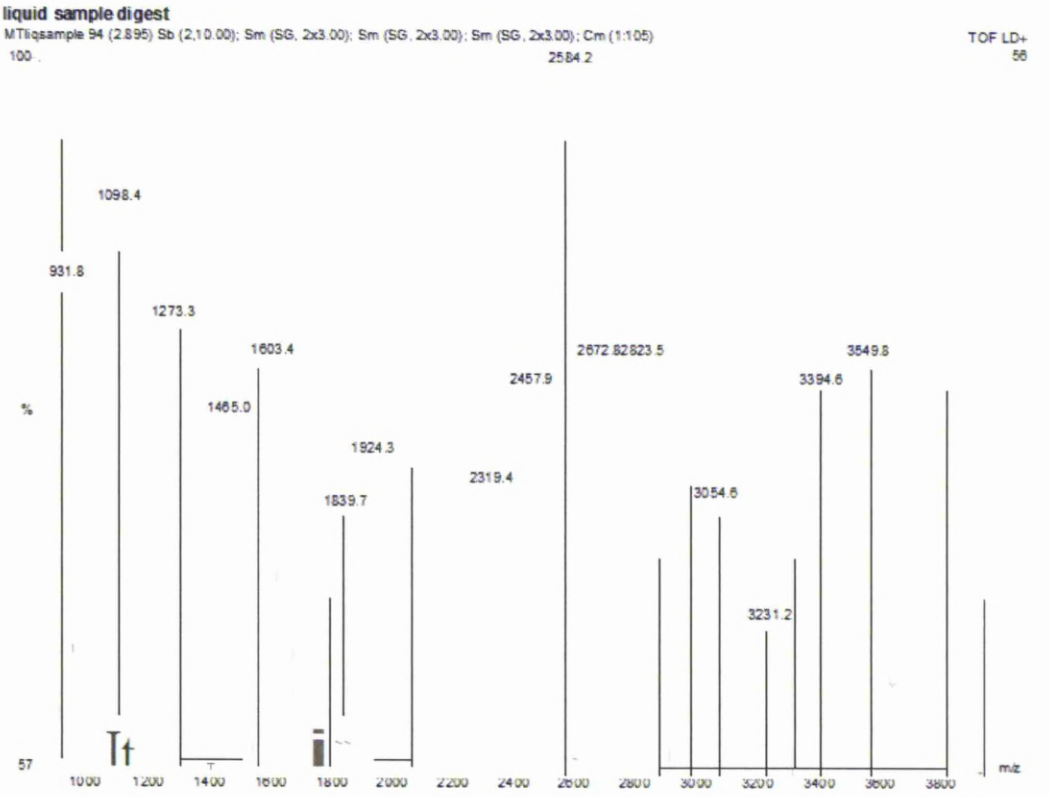
SAP97 PDZ 2



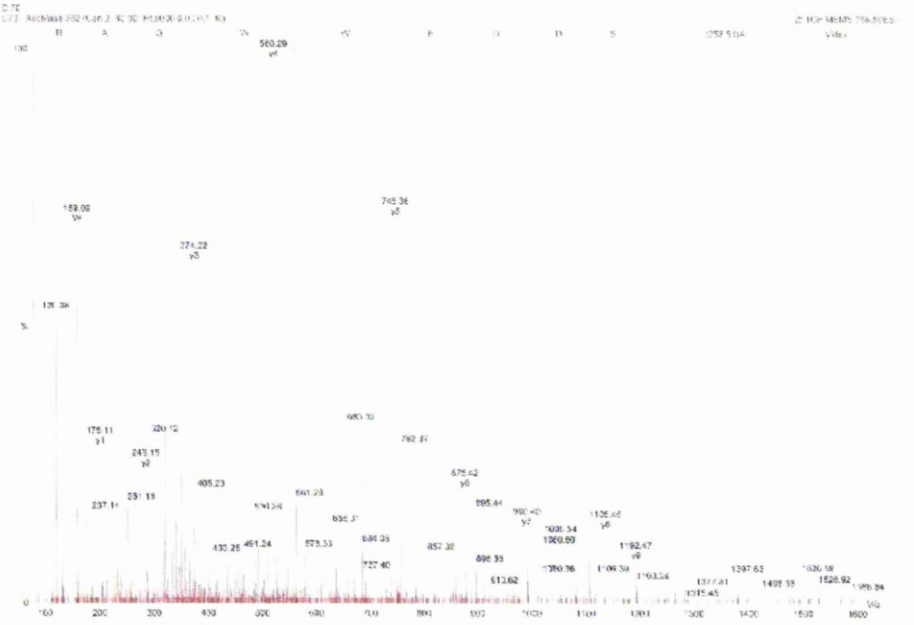
AKAP79 18-61


985.5
MT1 1001 (18.825)
TOF MS ES+
3.06e4
100
%
0
4597.1
4583.5
4582.1
4596.4
4179.3
4179.6
4455.1
4441.0
4605.8
4456.4
4463.6
4177.9
4468.9
4579.3
4618.5
4748.3
4243.6
4242.3
4569.3
4618.9
4748.9
4297.1
4568.7
4662.1
4934.5
4953.1
4313.1
4439.4
4844.5
4860.1
4174.5
4465.3
4721.4
4890.2
4313.8
4506.2
4508.3
4761.3
4987.5
4200
4300
4400
4500
4600
4700
4800
4900
mass

AKAP19-61 cleaved



D70 tryptic digest



NCBI Blast::lcl|27405 (10 letters)

Distance tree of results NEW

Sequences producing significant alignments:	Score (Bits)	E Value	
gb|EAX05336.1|	disco, large homolog 3 (neuroendocrine-dlg, Dr...		0.002	G
gb|EAW75093.1|	disco, large homolog 2, chapsyn-110 (Drosophil...		0.002	G
gb|EAW75095.1|	disco, large homolog 2, chapsyn-110 (Drosophil...		0.002	G
ref|NP_001070.2|	disco, large homolog 1 isoform 2 [Homo sapie...		0.002	UG
ref|NP_001091894.1|	disco, large homolog 1 isoform 1 [Homo sa...		0.002	UG
ref|NP_001355.2|	chapsyn-110 [Homo sapiens] >gb|EAW75094.1| d...		0.002	UG
dbj|BAC92777.1|	Presynaptic protein SAP97 variant [Homo sapiens]		0.002	G
dbj|BAC92565.1|	synapse-associated protein 102 variant [Homo ...		0.002	G
gb|CAI43973.1|	hypothetical protein [Homo sapiens]		0.002	G
ref|NP_065781.1|	synapse associated protein 102 isoform b [Ho...		0.002	UG

Alignments

>gb|EAX05336.1| G disco, large homolog 3 (neuroendocrine-dlg, Drosophila), isoform CRA_c [Homo sapiens]
Length=935

GENE ID: 1741 DLG3 disco, large homolog 3 (neuroendocrine-dlg, Drosophila) [Homo sapiens] (Over 10 PubMed links)

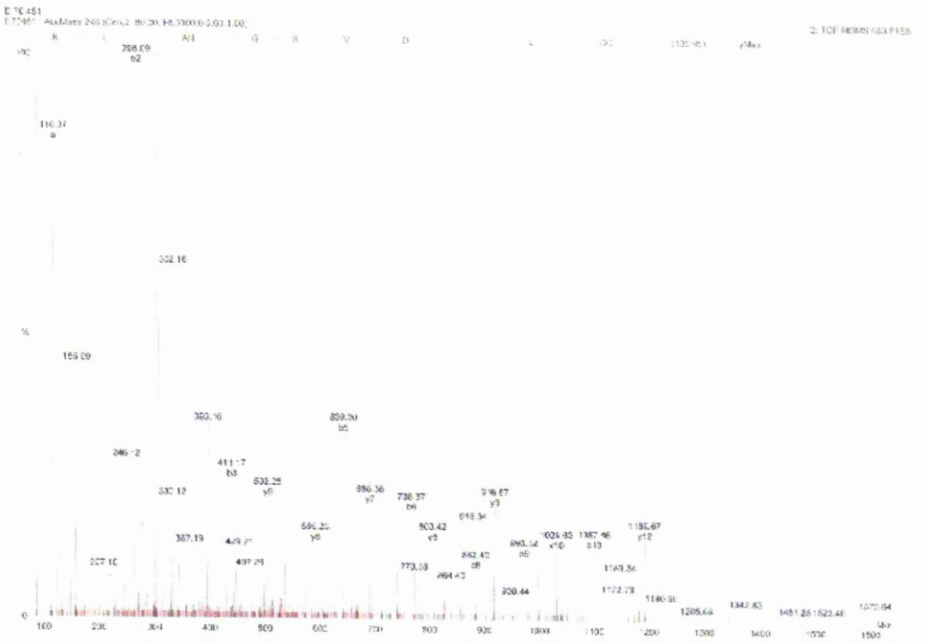
Score = 38.8 bits (84), Expect = 0.002
Identities = 10/10 (100%), Positives = 10/10 (100%), Gaps = 0/10 (0%)

Query 1 ASDDEAXQAR 10
ASDDEAXQAR
Sbjct 556 ASDDEAXQAR 565

>gb|EAW75093.1| G disco, large homolog 2, chapsyn-110 (Drosophila), isoform CRA_a [Homo sapiens]
Length=884

GENE ID: 1740 DLG2 disco, large homolog 2, chapsyn-110 (Drosophila) [Homo sapiens] (Over 10 PubMed links)

D461 tryptic digest



NCBI Blast:lel25310 (12 letters)

Distance tree of results NEW

Sequences producing significant alignments:	Score (Bits)	E Value	
gb|EAX05336.1|	disc, large homolog 3 (neuroendocrine-dlg, Dr...	37.5	0.004	G
gb|EAW75093.1|	disc, large homolog 2, chapsyn-110 (Drosophil...	37.5	0.004	G
gb|EAW77337.1|	disc, large homolog 2, chapsyn-110 (Drosophil...	37.5	0.004	G
ref|NP_004078.3|	disc, large homolog 1 isoform 2 [Homo sapie...	37.5	0.004	UG
ref|NP_001091894.1|	disc, large homolog 1 isoform 1 [Homo sa...	37.5	0.004	UG
ref|NP_001355.2|	chapsyn-110 [Homo sapiens] >gb|EAW75094.1| d...	37.5	0.004	UG
dbj|BAD96773.1|	Presynaptic protein SAP97 variant [Homo sapiens]	37.5	0.004	G
dbj|BAD96765.1|	synapse-associated protein 102 variant [Homo ...	37.5	0.004	G
ref|NP_065711.1|	synapse-associated protein 102 isoform b [Ho...	37.5	0.004	UG
sp|Q15700|DLG2_HUMAN	Disks large homolog 2 (Postsynaptic dens...	37.5	0.004	G

Alignments

>gb|EAX05336.1| G discs, large homolog 3 (neuroendocrine-dlg, Drosophila), isoform CRA_a [Homo sapiens]
Length=835

GENE ID: 1741 DLG3 discs, large homolog 3 (neuroendocrine-dlg, Drosophila) [Homo sapiens] (Over 10 PubMed links)

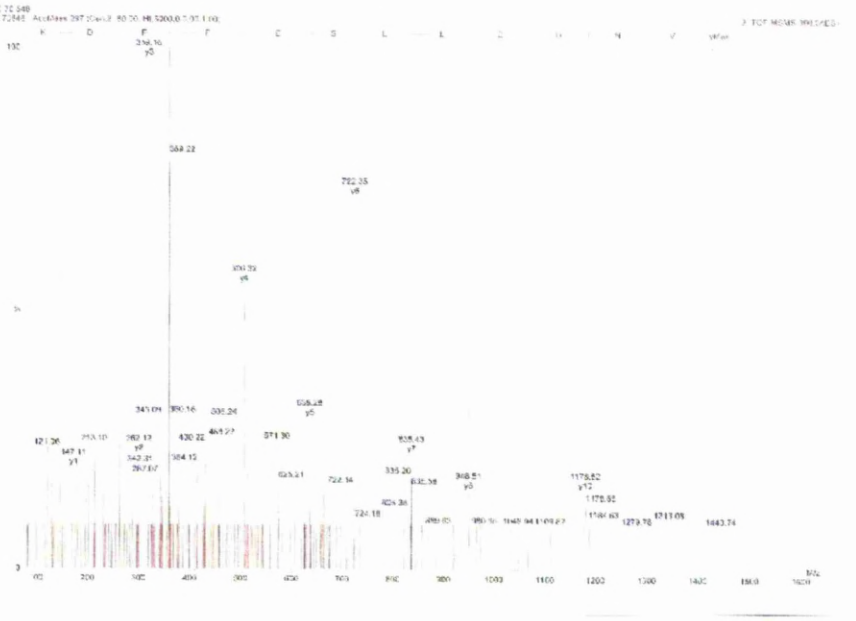
Score = 37.5 bits (85), Expect = 0.004
Identities = 11/11 (100%), Positives = 11/11 (100%), Gaps = 0/11 (0%)

Query 2 SSLDVSSKACK 12
SSLDVSSKACK
Sbjct 737 SSLDVSSKACK 747

>gb|EAW75093.1| G discs, large homolog 2, chapsyn-110 (Drosophila), isoform CRA_a [Homo sapiens]
Length=884

GENE ID: 1740 DLG2 discs, large homolog 2, chapsyn-110 (Drosophila)

D546 tryptic digest



NCBI Blast:ldl|31991 (12 letters)

Sequences producing significant alignments:

		Score (Bits)	E Value	
ref XP_001340947.2	PREDICTED: similar to discs, large (Droso...	41.9	0.006	G
gb AAI55673.1	Dlgh4 protein [Xenopus tropicalis]	41.9	0.006	G
gb AAH85435.1	Dlgh4 protein [Xenopus tropicalis]	41.9	0.006	G
gb AAI63063.1	Dlgh4 protein [Xenopus tropicalis]	41.9	0.006	G
ref NP_998863.1	discs, large (Drosophila) homolog 4 [Danio r...	41.6	0.006	UG
emb CAE00381.1	unnamed protein product [Tetraodon nigroviridis]	39.7	0.026	
ref XP_001370647.1	PREDICTED: similar to postsynaptic densit...	38.9	0.048	UG
gb EAX30233.1	discs, large homolog 4 (Drosophila), isoform C...	38.9	0.048	G
gb EAX30232.1	discs, large homolog 4 (Drosophila), isoform C...	38.9	0.048	G
ref NP_001103022.1	post-synaptic density protein 95 isoform ...	38.9	0.048	UG

Alignments

>ref|XP_001340947.2| G PREDICTED: similar to discs, large (Drosophila) homolog 4 (Danio rerio)
Length=776

Score = 41.9 bits (91), Expect = 0.006
Identities = 12/12 (100%), Positives = 12/12 (100%), Gaps = 0/12 (0%)

Query 1 VNIDLLSEFPDK 12
VNIDLLSEFPDK
Sbjct 597 VNIDLLSEFPDK 608

>gb|AAI55673.1| G Dlgh4 protein [Xenopus tropicalis]
Length=783

GENE ID: 779559 Dlgh4 postsynaptic density protein 95 [Xenopus tropicalis]
(10 or fewer PubMed links)

Score = 41.9 bits (91), Expect = 0.006
Identities = 12/12 (100%), Positives = 12/12 (100%), Gaps = 0/12 (0%)

Query 1 VNIDLLSEFPDK 12
VNIDLLSEFPDK